Association for Women in Mathematics Series

Hélène Barcelo
Gizem Karaali
Rosa Orellana *Editors*

Recent Trends in Algebraic Combinatorics

AWM
ASSOCIATION FOR WOMEN IN MATHEMATICS

Springer

Association for Women in Mathematics Series

Volume 16

Series editor
Kristin Lauter
Microsoft Research
Redmond, WA, USA

Association for Women in Mathematics Series

Focusing on the groundbreaking work of women in mathematics past, present, and future, Springer's Association for Women in Mathematics Series presents the latest research and proceedings of conferences worldwide organized by the Association for Women in Mathematics (AWM). All works are peer-reviewed to meet the highest standards of scientific literature, while presenting topics at the cutting edge of pure and applied mathematics. Since its inception in 1971, The Association for Women in Mathematics has been a non-profit organization designed to help encourage women and girls to study and pursue active careers in mathematics and the mathematical sciences and to promote equal opportunity and equal treatment of women and girls in the mathematical sciences. Currently, the organization represents more than 3000 members and 200 institutions constituting a broad spectrum of the mathematical community, in the United States and around the world.

More information about this series at http://www.springer.com/series/13764

Hélène Barcelo · Gizem Karaali
Rosa Orellana
Editors

Recent Trends in Algebraic Combinatorics

Springer

Editors
Hélène Barcelo
Mathematical Sciences Research Institute
Berkeley, CA, USA

Rosa Orellana
Department of Mathematics
Dartmouth College
Hanover, NH, USA

Gizem Karaali
Department of Mathematics
Pomona College
Claremont, CA, USA

ISSN 2364-5733 ISSN 2364-5741 (electronic)
Association for Women in Mathematics Series
ISBN 978-3-030-05140-2 ISBN 978-3-030-05141-9 (eBook)
https://doi.org/10.1007/978-3-030-05141-9

Library of Congress Control Number: 2018963041

Mathematics Subject Classification (2010): 05Exx, 16-xx, 20-xx

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book is a collection of survey articles on some of the exciting recent developments in algebraic combinatorics. It also contains a tutorial on Schubert calculus, as we felt that such an article would constitute an important addition to the literature.

Algebraic combinatorics is a vast area of research, and attempting to be exhaustive in our coverage would have been a mistake. Rather (after reaching out to many experts in the field for advice), we chose a small number of topics that are currently enjoying broad interest and rapid growth. And after many thoughtful discussions, we sought out the authors of this volume—several writers known for their expository skills as well as a few junior researchers with active research programs who were willing to share their knowledge. Writing a good survey is a deceptively tricky endeavor, as decisions on what to include (or not) and how to best present it lurk at the corner of every sentence! Nonetheless, our contributors did a remarkable job of introducing exciting directions of current research in algebraic combinatorics from its foundational questions to the boundaries of what is known today.

More specifically, the book contains four surveys focusing on representation theory, symmetric functions, invariant theory, and the combinatorics of Young tableaux. The other five surveys address subjects at the intersection of algebra, combinatorics, and geometry: the study of polytopes, lattice points, hyperplane arrangements, crystal graphs, and Grassmannians. The surveys are written at an introductory level that emphasizes recent developments and open problems. The tutorial on Schubert calculus is written in an interactive way and is intended as a guide for combinatorialists wishing to understand and appreciate the geometric and topological aspects of Schubert calculus, as well as for geometric-minded researchers seeking to gain familiarity with the relevant combinatorial tools in this area.

Each article in this volume was reviewed independently by two referees, and we are simply amazed by, and deeply grateful for, the generosity of those referees. Their care in reviewing the articles and their constructive and judicious suggestions have been invaluable. The authors of this volume wish to warmly thank the referees

for their contributions, and we three editors in turn offer our heartfelt thanks to both the authors and the referees. We are very proud to have edited such a fine book.

We hope that you will enjoy reading these articles as much as we have, and (as one referee wrote) you will find that the stories flow well and are inspiring!

Berkeley, USA Hélène Barcelo
Claremont, USA Gizem Karaali
Hanover, USA Rosa Orellana

Contents

Partition Algebras and the Invariant Theory of the Symmetric Group

Georgia Benkart and Tom Halverson

Abstract The symmetric group S_n and the partition algebra $\mathsf{P}_k(n)$ centralize one another in their actions on the k-fold tensor power $\mathsf{M}_n^{\otimes k}$ of the n-dimensional permutation module M_n of S_n. The duality afforded by the commuting actions determines an algebra homomorphism $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ from the partition algebra to the centralizer algebra $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$, which is a surjection for all $k, n \in \mathbb{Z}_{\geq 1}$, and an isomorphism when $n \geq 2k$. We present results that can be derived from the duality between S_n and $\mathsf{P}_k(n)$, for example, (i) expressions for the multiplicities of the irreducible S_n-summands of $\mathsf{M}_n^{\otimes k}$, (ii) formulas for the dimensions of the irreducible modules for the centralizer algebra $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$, (iii) a bijection between vacillating tableaux and set-partition tableaux, (iv) identities relating Stirling numbers of the second kind and the number of fixed points of permutations, and (v) character values for the partition algebra $\mathsf{P}_k(n)$. When $2k > n$, the map $\Phi_{k,n}$ has a nontrivial kernel which is generated as a two-sided ideal by a single idempotent. We describe the kernel and image of $\Phi_{k,n}$ in terms of the orbit basis of $\mathsf{P}_k(n)$ and explain how the surjection $\Phi_{k,n}$ can also be used to obtain the fundamental theorems of invariant theory for the symmetric group.

Keywords Symmetric group · Partition algebra · Schur–Weyl duality
Invariant theory

Mathematics Subject Classification (2010) MSC 05E10 · MSC 20C30

The second author gratefully acknowledges partial support from Simons Foundation grant 283311.

G. Benkart (✉)
University of Wisconsin - Madison, Madison, WI 53706, USA
e-mail: benkart@math.wisc.edu

T. Halverson
Macalester College, Saint Paul, MN 55105, USA
e-mail: halverson@macalester.edu

H. Barcelo et al. (eds.), *Recent Trends in Algebraic Combinatorics*,
Association for Women in Mathematics Series 16,
https://doi.org/10.1007/978-3-030-05141-9_1

1 Introduction

Throughout we assume $\mathbb{F}$ is a field of characteristic zero. The symmetric group S_n has a natural action on its n-dimensional permutation module M_n over $\mathbb{F}$ by permuting the basis elements. The focus of this article is on tensor powers $\mathsf{M}_n^{\otimes k}$ of M_n, which are S_n-modules under the diagonal action (see (1.2)). For $n, k \in \mathbb{Z}_{\geq 1}$, the partition algebra $\mathsf{P}_k(n)$ is an associative $\mathbb{F}$-algebra with basis indexed by the set partitions of $\{1, 2, \ldots, 2k\}$ and with multiplication given by concatenation of set-partition diagrams (described in Sect. 4). In [22], V. Jones constructed a surjective algebra homomorphism

$$\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k}) \tag{1.1}$$

from the partition algebra onto the centralizer algebra $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ of transformations that commute with the action of S_n on $\mathsf{M}_n^{\otimes k}$. When $n \geq 2k$, this surjection is an isomorphism.

Schur–Weyl duality relates the representation theory of a group and its centralizer algebra acting simultaneously on a tensor power representation (see, e.g., [16, Chap. 9] or [19, Sects. 3 and 5]). The classical case considers the general linear group $\mathsf{GL_n}$ and its fundamental module $\mathsf{M}_n = \mathbb{F}^n$. The centralizer of $\mathsf{GL_n}$ on the tensor power $\mathsf{M}_n^{\otimes k}$ is provided by the surjection $\mathbb{F}\mathsf{S}_k \to \mathsf{End}_{\mathsf{GL}_n}(\mathsf{M}_n^{\otimes k})$, where S_k acts by permuting the tensor factors. This map is an algebra isomorphism when $n \geq k$. Idempotents in the group algebra $\mathbb{F}\mathsf{S}_k$ determine the projection maps onto the irreducible GL_n-summands of $\mathsf{M}_n^{\otimes k}$, and this enabled Schur to construct all the irreducible polynomial representations of GL_n from the tensor powers $\mathsf{M}_n^{\otimes k}$. A similar surjective algebra homomorphism, $\mathsf{B}_k(n) \to \mathsf{End}_{\mathsf{O}_n}(\mathsf{M}_n^{\otimes k})$ from the Brauer algebra $\mathsf{B}_k(n)$ to the algebra $\mathsf{End}_{\mathsf{O}_n}(\mathsf{M}_n^{\otimes k})$ of transformations of $\mathsf{M}_n^{\otimes k}$ commuting with the action of the orthogonal group O_n, is an isomorphism when $n \geq k$ and has been used to study the action of O_n on $\mathsf{M}_n^{\otimes k}$. When restricted to the subgroup of permutation matrices in O_n, the module M_n becomes the permutation module of S_n, and the inclusions $\mathsf{S}_n \subset \mathsf{O}_n \subset \mathsf{GL}_n$ imply reverse inclusions of their centralizer algebras. In particular, when $n \geq 2k$, the centralizer chain is $\mathsf{P}_k(n) \supset \mathsf{B}_k(n) \supset \mathbb{F}\mathsf{S}_k$. In this duality picture, there are two different symmetric groups acting on $\mathsf{M}_n^{\otimes k}$, namely the group S_n which acts diagonally and the group S_k which acts as tensor place permutations.

The Schur–Weyl duality afforded by the homomorphism $\Phi_{k,n}$ in (1.1) between the partition algebra $\mathsf{P}_k(n)$ and the symmetric group S_n in their commuting actions on $\mathsf{M}_n^{\otimes k}$ enables information to flow back and forth between S_n and $\mathsf{P}_k(n)$. Indeed, the symmetric group S_n has been used to

- develop the combinatorial representation theory of the partition algebras $\mathsf{P}_k(n)$ as k varies [3, 4, 13, 14, 17–19, 28, 29] and
- study the Potts lattice model of interacting spins in statistical mechanics [26–28];

while the partition algebra $\mathsf{P}_k(n)$ has been used to

- study eigenvalues of random permutation matrices [14],

- prove results about Kronecker coefficients for S_n-modules [6–8],
- investigate the centralizer algebras of the binary tetrahedral, octahedral, and icosahedral subgroups of the special unitary group SU_2 acting on tensor powers of $\mathsf{V} = \mathbb{C}^2$ via the McKay correspondence [1], and
- construct a nonhomogeneous basis for the ring of symmetric functions and show that the irreducible characters of S_n can be obtained by evaluating this basis on the eigenvalues of the permutation matrices, which has led to results on reduced Kronecker coefficients [31].

The irreducible modules S_n^λ for the symmetric group S_n are indexed by partitions λ of n, indicated here by $\lambda \vdash n$. The n-dimensional permutation module M_n for S_n has basis elements $\mathsf{v}_1, \mathsf{v}_2, \ldots, \mathsf{v}_n$, which are permuted by the elements σ of S_n, $\sigma.\mathsf{v}_i = \mathsf{v}_{\sigma(i)}$. The module M_n has a decomposition into irreducible S_n-modules, $\mathsf{M}_n = \mathsf{S}_n^{[n]} \oplus \mathsf{S}_n^{[n-1,1]}$, where the vector $\mathsf{v}_1 + \mathsf{v}_2 + \cdots + \mathsf{v}_n$ spans a copy of the trivial S_n-module $\mathsf{S}_n^{[n]}$, and the vectors $\mathsf{v}_i - \mathsf{v}_{i+1}$ for $1 \le i \le n-1$ form a basis for a copy of the $(n-1)$-dimensional module $\mathsf{S}_n^{[n-1,1]}$. The module $\mathsf{S}_n^{[n-1,1]}$ is often referred to as the “reflection representation,” since the transposition switching i and j is a reflection about the hyperplane orthogonal to $\mathsf{v}_i - \mathsf{v}_j$. The S_n-action on M_n extends to give the tensor power $\mathsf{M}_n^{\otimes k}$ the structure of an S_n-module:

$$\sigma.\left(\mathsf{v}_{i_1} \otimes \mathsf{v}_{i_2} \otimes \cdots \otimes v_{i_k}\right) = \sigma.\mathsf{v}_{i_1} \otimes \sigma.\mathsf{v}_{i_2} \otimes \cdots \otimes \sigma.\mathsf{v}_{i_k} = \mathsf{v}_{\sigma(i_1)} \otimes \mathsf{v}_{\sigma(i_2)} \otimes \cdots \otimes \mathsf{v}_{\sigma(i_k)}. \tag{1.2}$$

In Sects. 2 and 3, we describe two different ways to determine the multiplicity $\mathsf{m}_{k,n}^\lambda$ of S_n^λ in the decomposition

$$\mathsf{M}_n^{\otimes k} = \bigoplus_{\lambda \vdash n} \mathsf{m}_{k,n}^\lambda \, \mathsf{S}_n^\lambda$$

of $\mathsf{M}_n^{\otimes k}$ into irreducible S_n-summands. The first approach, using restriction and induction on the pair $(\mathsf{S}_n, \mathsf{S}_{n-1})$, leads naturally to a bijection between the set of vacillating k-tableaux of shape λ and the set of paths in the Bratteli diagram $\mathcal{B}(\mathsf{S}_n, \mathsf{S}_{n-1})$ from the partition $[n]$ at the top of $\mathcal{B}(\mathsf{S}_n, \mathsf{S}_{n-1})$ to λ at level k. Both sets have cardinality $\mathsf{m}_{k,n}^\lambda$. The second way, adopted from [4], uses permutation modules for S_n and results in an exact expression for the multiplicity of an irreducible S_n-summand of $\mathsf{M}_n^{\otimes k}$. In particular, we describe how [4, Thm. 5.5] implies that

$$\mathsf{m}_{k,n}^\lambda = \sum_{t=|\lambda^\#|}^{n} \left\{ {k \atop t} \right\} f^{\lambda/[n-t]},$$

where the *Stirling number of the second kind* $\left\{ {k \atop t} \right\}$ counts the number of ways to partition a set of k objects into t nonempty disjoint blocks (subsets), $f^{\lambda/[n-t]}$ is the number of standard tableaux of skew shape $\lambda/[n-t]$, and $\lambda^\#$ is the partition obtained from λ by removing its first (largest) part. Therefore,

$$\mathsf{m}_{k,n}^{\lambda} = \left| \left\{ (\pi, \mathsf{S}) \;\middle|\; \begin{array}{l} \pi \text{ is a set partition of } \{1, 2, \ldots, k\} \text{ into } t \text{ parts, where } |\lambda^{\#}| \le t \le n \\ \mathsf{S} \text{ is a standard tableau of skew shape } \lambda/[n-t] \end{array} \right\} \right|.$$

In Sect. 3.2, we consider set-partition tableaux (tableaux whose boxes are filled with sets of numbers from $\{0, 1, \ldots, k\}$) and demonstrate a bijection between set-partition tableaux and vacillating tableaux. Different bijections between these sets are given in [10, 18]. However, the bijections in those papers apply only when $n \ge 2k$. The bijection here has the advantage of working for all $k, n \in \mathbb{Z}_{\ge 1}$. Such set-partition tableaux also appear in the recent investigations of Orellana and Zabrocki [32, Example 2].

At the core of these results is the duality between the representation theories of the symmetric group S_n and the centralizer algebra,

$$\mathsf{Z}_{k,n} := \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k}) = \{\varphi \in \mathsf{End}(\mathsf{M}_n^{\otimes k}) \mid \varphi\sigma(x) = \sigma\varphi(x), \ \ \sigma \in \mathsf{S}_n, x \in \mathsf{M}_n^{\otimes k}\}, \tag{1.3}$$

of its action on $\mathsf{M}_n^{\otimes k}$. Schur–Weyl duality tells us that the irreducible modules $\mathsf{Z}_{k,n}^{\lambda}$ for the semisimple associative algebra $\mathsf{Z}_{k,n}$ are indexed by the subset Λ_{k,S_n} of partitions λ of n such that S_n^{λ} occurs with multiplicity at least one in the decomposition of $\mathsf{M}_n^{\otimes k}$ into irreducible S_n-summands. Moreover,

- $$\mathsf{M}_n^{\otimes k} \cong \underbrace{\bigoplus_{\lambda \in \Lambda_{k,\mathsf{S}_n}} \mathsf{m}_{k,n}^{\lambda} \mathsf{S}_n^{\lambda}}_{\text{as an } \mathsf{S}_n\text{-module}} \cong \underbrace{\bigoplus_{\lambda \in \Lambda_{k,\mathsf{S}_n}} f^{\lambda} \mathsf{Z}_{k,n}^{\lambda}}_{\text{as a } \mathsf{Z}_{k,n}\text{-module}}; \tag{1.4}$$
- $\dim(\mathsf{Z}_{k,n}^{\lambda}) = \mathsf{m}_{k,n}^{\lambda}$ (the multiplicity of S_n^{λ} in $\mathsf{M}_n^{\otimes k}$); (1.5)
- $\mathsf{mult}(\mathsf{Z}_{k,n}^{\lambda}) = \dim(\mathsf{S}_n^{\lambda}) = f^{\lambda}$ (the number of standard tableaux of shape λ); (1.6)
- $$\dim(\mathsf{Z}_{k,n}) = \sum_{\lambda \in \Lambda_{k,\mathsf{S}_n}} (\mathsf{m}_{k,n}^{\lambda})^2 = \mathsf{m}_{2k,n}^{[n]} = \dim\left(\mathsf{Z}_{2k,n}^{[n]}\right). \tag{1.7}$$

The first equality in (1.7) is a consequence of (1.5) and Artin–Wedderburn theory, since $\dim(\mathsf{Z}_{k,n})$ is the sum of the squares of the dimensions of its irreducible modules $\mathsf{Z}_{k,n}^{\lambda}$. The second equality in (1.7) comes from the isomorphism $\mathsf{Z}_{k,n} = \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k}) \cong (\mathsf{M}_n^{\otimes 2k})^{\mathsf{S}_n}$, since the S_n-invariants in $\mathsf{M}_n^{\otimes 2k}$ correspond to copies of the trivial one-dimensional module $\mathsf{S}_n^{[n]}$ indexed by the one-part partition $[n]$, and $\mathsf{m}_{2k,n}^{[n]} = \dim\left(\mathsf{Z}_{2k,n}^{[n]}\right)$ is the number of trivial summands of $\mathsf{M}_n^{\otimes 2k}$.

Three additional Schur–Weyl duality results can be found in [5, Cor. 2.5]. From those results, we know that for any finite group G and any finite-dimensional G-module V, the dimension of the space $\left(\mathsf{V}^{\otimes k}\right)^{\mathsf{G}}$ of G-invariants is the average of the character values $\chi_{\mathsf{V}^{\otimes k}}(g) = \chi_{\mathsf{V}}(g)^k$ as g ranges over the elements of G (compare also [15, Sect. 2.2]). When V is isomorphic to its dual G-module, the dimension of the centralizer algebra $\mathsf{End}_{\mathsf{G}}(\mathsf{V}^{\otimes k})$ is the average of the character values

$\chi_{V^{\otimes 2k}}(g) = \chi_V(g)^{2k}$. Specializing those results and part (i) of [5, Cor. 2.5] to the case $G = S_n$ and $V = M_n$ gives the following:

- $$\dim\left(\left(M_n^{\otimes k}\right)^{S_n}\right) = \frac{1}{n!}\sum_{\sigma\in S_n}\chi_{M_n}(\sigma)^k = \frac{1}{n!}\sum_{\sigma\in S_n}F(\sigma)^k; \tag{1.8}$$
- $$\dim(Z_{k,n}) = \dim(Z_{2k,n}^{[n]}) = \frac{1}{n!}\sum_{\sigma\in S_n}F(\sigma)^{2k}; \tag{1.9}$$
- $$\dim(Z_{k,n}^{\lambda}) = \frac{1}{n!}\sum_{\sigma\in S_n}\chi_{M_n}(\sigma)^k\ \chi_\lambda(\sigma^{-1}) = \frac{1}{n!}\sum_{\sigma\in S_n}F(\sigma)^k\ \chi_\lambda(\sigma), \tag{1.10}$$

where $\chi_{M_n}(\sigma) = F(\sigma)$ (the number of fixed points of σ) and χ_λ is the character of the irreducible S_n-module S_n^λ. In (1.10), we have used the fact that σ and σ^{-1} have the same cycle structures, hence the same character values.

In [14], Farina and Halverson develop the character theory of partition algebras and use partition algebra characters to prove (1.9). The arguments in [14], which are based on results from [17], require the assumption $n \geq 2k$ so that $Z_{k,n} \cong P_k(n)$. The results in (1.8)–(1.10) are true for all $k, n \geq 1$.

When V is a finite-dimensional module for a finite group G, the tensor power $V^{\otimes k}$ has a multiplicity-free decomposition

$$V^{\otimes k} = \bigoplus_\nu \left(G^\nu \otimes Z_{k,G}^\nu\right)$$

into irreducible summands as a bimodule for $G \times Z_{k,G}$, where $Z_{k,G} = \mathrm{End}_G(V^{\otimes k})$. Consequently, the characters of G, $Z_{k,G}$ and $G \times Z_{k,G}$ are intertwined by the following equation,

$$\psi_{V^{\otimes k}}(g \times z) = \sum_\nu \chi_\nu(g)\ \xi_\nu(z).$$

The irreducible characters of G are orthonormal with respect to the well-known inner product on class functions of G defined by $\langle\alpha, \beta\rangle = |G|^{-1}\sum_{g\in G}\alpha(g)\beta(g^{-1})$ (see, e.g., [15, Thm. 2.12] or [35, Thm. 1.93]). Therefore, since $\psi_{V^{\otimes k}}(\cdot \times z) : G \to \mathbb{F}$ is a class function on G for each $z \in Z_{k,G}$, we have

$$\langle\psi_{V^{\otimes k}}(\cdot \times z), \chi_\nu\rangle = \frac{1}{|G|}\sum_{g\in G}\psi_{V^{\otimes k}}(g \times z)\chi_\nu(g^{-1}) = \xi_\nu(z).$$

Thus, the commuting actions of G and $Z_{k,G}$ on $V^{\otimes k}$ lead to one further Schur–Weyl duality result, namely an expression for the irreducible characters ξ_ν of $Z_{k,G}$:

- $$\xi_\nu(z) = \frac{1}{|G|}\sum_{g\in G}\psi_{V^{\otimes k}}(g \times z)\chi_\nu(g^{-1}). \tag{1.11}$$

In Sect. 4.5, we explain how these ideas, when combined with results from [17], provide the formula in Theorem 4.17 for the characters of the partition algebra $\mathsf{P}_k(n)$.

The surjective algebra homomorphism $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{Z}_{k,n} = \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ enables us to study the S_n-module $\mathsf{M}_n^{\otimes k}$ using the partition algebra. Since the set partitions of the set $[1, 2k] := \{1, 2, \dots, 2k\}$ index an $\mathbb{F}$-basis for $\mathsf{P}_k(n)$, the image of these basis elements linearly spans the S_n-module endomorphisms $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$, and the generators of $\mathsf{P}_k(n)$ generate $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ as an associative algebra.

We let Π_{2k} be the set of set partitions of $[1, 2k]$; for example, $\big\{1, 7, 8, 10 \,|\, 2, 5 \,|\, 4, 9, 11 \,|\, 3, 6, 12, 14 \,|\, 13\big\}$ is a set partition in Π_{14} with 5 blocks (subsets). The number of set partitions in Π_{2k} with t blocks is given by the Stirling number of the second kind $\left\{ {2k \atop t} \right\}$, and thus $\mathsf{P}_k(n)$ has dimension equal to the *Bell number* $\mathsf{B}(2k) = \sum_{t=1}^{2k} \left\{ {2k \atop t} \right\}$.

The algebra $\mathsf{P}_k(n)$ has two distinguished bases—the diagram basis $\{d_\pi \mid \pi \in \Pi_{2k}\}$ and the orbit basis $\{x_\pi \mid \pi \in \Pi_{2k}\}$. We describe the change of basis matrix between the diagram basis and the orbit basis in terms of the Möbius function of the set-partition lattice (see Sect. 4.3). Section 4.4 is devoted to a description of multiplication in the orbit basis of $\mathsf{P}_k(n)$. The actions of the diagram basis element d_π and the orbit basis element x_π on $\mathsf{M}_n^{\otimes k}$ afforded by the representation $\Phi_{k,n}$ are given explicitly in Sect. 5.2 for any $\pi \in \Pi_{2k}$. The orbit basis is key to understanding the image and the kernel of $\Phi_{k,n}$, as we explain in Sect. 5.3. An analog of the orbit basis exists in a broader context, and in Sect. 4.2 we explain how to construct an orbit basis for the centralizer algebra of a tensor power of any permutation module for an arbitrary finite group G.

The centralizer algebra $\mathsf{Z}_{k,n} = \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ has a basis consisting of the images $\Phi_{k,n}(x_\pi)$, where π ranges over the set partitions in Π_{2k} with no more than n blocks. As a consequence of that result and (1.9), we have

$$\dim(\mathsf{Z}_{k,n}) = \sum_{t=1}^{n} \left\{ {2k \atop t} \right\} = \frac{1}{n!} \sum_{\sigma \in \mathsf{S}_n} \mathsf{F}(\sigma)^{2k}. \tag{1.12}$$

We prove additional relations between Stirling numbers of the second kind and fixed points of permutations in Sect. 3.1 (more specifically, see Theorem 3.7, Proposition 3.9, and Corollary 3.11).

The diagram basis elements d_π in $\mathsf{P}_{k+1}(n)$ corresponding to set partitions π having $k+1$ and $2(k+1)$ in the same block form a subalgebra $\mathsf{P}_{k+\frac{1}{2}}(n)$ of $\mathsf{P}_{k+1}(n)$ under diagram multiplication. Identifying $\mathsf{P}_k(n)$ with the span of the diagrams in $\mathsf{P}_{k+\frac{1}{2}}(n)$ that have a block consisting solely of the two elements $k+1, 2(k+1)$ gives a tower of algebras,

$$\mathbb{F} = \mathsf{P}_0(n) \cong \mathsf{P}_{\frac{1}{2}}(n) \subset \mathsf{P}_1(n) \subset \cdots \subset \mathsf{P}_k(n) \subset \mathsf{P}_{k+\frac{1}{2}}(n) \subset \mathsf{P}_{k+1}(n) \subset \cdots . \tag{1.13}$$

If we regard M_n as a module for the symmetric group S_{n-1} by restriction, where elements of S_{n-1} fix the last basis vector v_n, then there is a surjective algebra homomorphism

$$\Phi_{k+\frac{1}{2},n} : \mathsf{P}_{k+\frac{1}{2}}(n) \to \mathsf{End}_{\mathsf{S}_{n-1}}(\mathsf{M}_n^{\otimes k}), \tag{1.14}$$

which is an isomorphism when $n \geq 2k+1$. The intermediate algebras $\mathsf{P}_{k+\frac{1}{2}}(n)$ have proven useful in developing the structure and representation theory of partition algebras (see, e.g., [19, 29]), and they come into play here in the construction of vacillating k-tableaux. The centralizer algebras $\mathsf{End}_{\mathsf{S}_{n-1}}(\mathsf{M}_n^{\otimes k})$ are also closely tied to the restriction and induction functors that produce the Bratteli diagram $\mathcal{B}(\mathsf{S}_n, \mathsf{S}_{n-1})$ in Sect. 2.3.

For a group G and a finite-dimensional G-module V, the tensor product $\mathsf{V} \otimes \mathsf{V}^*$ of V with its dual module V^* is isomorphic as a G-module to $\mathsf{End}(\mathsf{V})$ via the mapping $\mathsf{v} \otimes \phi \mapsto A_{\mathsf{v},\phi}$, where $A_{\mathsf{v},\phi}(\mathsf{u}) = \phi(\mathsf{u})\mathsf{v}$. Since $g.(\mathsf{v} \otimes \phi) = g.\mathsf{v} \otimes g.\phi$, where $(g.\phi)(\mathsf{u}) = \phi(g^{-1}.\mathsf{u})$, and $g.A = gAg^{-1}$ as transformations on V for all $A \in \mathsf{End}(\mathsf{V})$, we have

$$g.(\mathsf{v} \otimes \phi) = \mathsf{v} \otimes \phi \iff gA_{\mathsf{v},\phi}g^{-1} = A_{\mathsf{v},\phi} \iff gA_{\mathsf{v},\phi} = A_{\mathsf{v},\phi}g \iff A_{\mathsf{v},\phi} \in \mathsf{End}_{\mathsf{G}}(\mathsf{V}).$$

Thus, the space of G-invariants $(\mathsf{V} \otimes \mathsf{V}^*)^{\mathsf{G}} = \{\mathsf{v} \otimes \phi \mid g.(\mathsf{v} \otimes \phi) = \mathsf{v} \otimes \phi\}$ is isomorphic to the centralizer algebra $\mathsf{End}_{\mathsf{G}}(\mathsf{V})$.

In the particular case of the symmetric group S_n and its permutation module M_n, we can identify the centralizer algebra $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ with the space

$$(\mathsf{M}_n^{\otimes 2k})^{\mathsf{S}_n} \cong \left(\mathsf{M}_n^{\otimes k} \otimes (\mathsf{M}_n^{\otimes k})^*\right)^{\mathsf{S}_n}$$

of S_n-invariants, as M_n is isomorphic to its dual S_n-module M_n^* (this was used in (1.7) and (1.9)). The fact that $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ is a surjection tells us that $\mathsf{P}_k(n)$ generates all of the S_n-tensor invariants $(\mathsf{M}_n^{\otimes 2k})^{\mathsf{S}_n}$ and provides the First Fundamental Theorem of Invariant Theory for S_n (see Theorem 6.1 originally proven by Jones [22]). When $2k > n$, the surjection $\Phi_{k,n}$ has a nontrivial kernel. As $2k$ increases in relation to n, the kernel becomes quite significant, and $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ is only a shadow of the full partition algebra (This is illustrated in the table of dimensions in Fig. 5.). We have shown in [3] that when $2k > n$, the kernel of $\Phi_{k,n}$ is generated as a two-sided ideal by a single essential idempotent $\mathsf{e}_{k,n}$ (see (5.20)). Identifying $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k}) \cong \left(\mathsf{M}_n^{\otimes 2k}\right)^{\mathsf{S}_n}$ with $\mathsf{P}_k(n)/\mathsf{ker}\,\Phi_{k,n}$ $(= \mathsf{P}_k(n)/\langle \mathsf{e}_{k,n}\rangle$ when $2k > n)$, we have the following.

Theorem 1.15 [3, Thm. 5.19] *(Second Fundamental Theorem of Invariant Theory for S_n) For all $k, n \in \mathbb{Z}_{\geq 1}$, $\mathsf{im}\,\Phi_{k,n} = \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ is generated by the partition algebra generators and relations in Theorem 6.5(a)–(c) together with the one additional relation $\mathsf{e}_{k,n} = 0$ in the case that $2k > n$. When $k \geq n$, the relation $\mathsf{e}_{k,n} = 0$ can be replaced with $\mathsf{e}_{n,n} = 0$.*

Classical Schur–Weyl duality provides an analogous result for the general linear group GL_n and its action by matrix multiplication on the space $\mathsf{V} = \mathsf{M}_n = \mathbb{F}^n$ of $n \times 1$ column vectors (more details can be found in [16, Sect. 9.1]). The surjection $\mathbb{F}\mathsf{S}_k \to \mathsf{End}_{\mathsf{GL}_n}(\mathsf{V}^{\otimes k})$, given by the place permutation action of S_k on the tensor factors of V,

is an isomorphism if $n \geq k$, and in that case, $\mathbb{F}\mathsf{S}_k \cong \left(\mathsf{V}^{\otimes k} \otimes (\mathsf{V}^{\otimes k})^*\right)^{\mathsf{GL}_n}$. When $k \geq n+1$, the kernel is generated by a single essential idempotent $\sum_{\sigma \in \mathsf{S}_{n+1}} (-1)^{\mathsf{sgn}(\sigma)} \sigma$ in the group algebra $\mathbb{F}\mathsf{S}_k$. Thus, the second fundamental theorem comes from the standard presentation by generators and relations for the symmetric group S_k by imposing the additional relation $\sum_{\sigma \in \mathsf{S}_{n+1}} (-1)^{\mathsf{sgn}(\sigma)} \sigma = 0$ when $k \geq n+1$.

In [9], Brauer introduced algebras $\mathsf{B}_k(n)$, now known as *Brauer algebras*, that centralize the action of the orthogonal group O_n and symplectic group Sp_{2n} on tensor powers of their defining modules. The surjective algebra homomorphisms $\mathsf{B}_k(n) \to \mathsf{End}_{\mathsf{O}_n}(\mathsf{V}^{\otimes k})$ and $\mathsf{B}_k(-2n) \to \mathsf{End}_{\mathsf{Sp}_{2n}}(\mathsf{V}^{\otimes k})$ (where $\mathsf{V} = \mathsf{M}_n$ for O_n and $\mathsf{V} = \mathsf{M}_{2n}$ for Sp_{2n}) defined in [9] provide the First Fundamental Theorem of Invariant Theory for these groups (An exposition of these results appears in [16, Sect. 4.3.2].). Generators for the kernels of these surjections give the Second Fundamental Theorem of Invariant Theory for the orthogonal and symplectic groups. In [20], Hu and Xiao showed that the kernel of the surjection in the symplectic case is principally generated when $k \geq n+1$. Lehrer and Zhang [23, 24] identified a generator for the kernel in both the symplectic and orthogonal cases. The Brauer diagram category is an essential ingredient in these papers (and also in the symplectic group investigations of Rubey and Westbury [33, 34]) for proving that the kernel of the surjection is principally generated and in establishing analogous results for the quantum analog of the Brauer algebra, the Birman–Murakami–Wenzl algebra. As shown in [24, Lemma 5.2] (compare [33, Sects. 7.4 and 7.6] and [34] which establish connections with cyclic sieving), the element $E = \frac{1}{(n+1)!} \sum_d d$, obtained by summing all the Brauer diagrams d with $2n+2$ vertices, is a central idempotent in $\mathsf{B}_{n+1}(-2n)$ corresponding to the one-dimensional trivial $\mathsf{B}_{n+1}(-2n)$-module, and it generates the kernel of the surjection $\mathsf{B}_k(-2n) \to \mathsf{End}_{\mathsf{Sp}_{2n}}(\mathsf{V}^{\otimes k})$ for all $k \geq n+1$. As a result, the fundamental theorems of invariant theory for the symplectic and orthogonal groups can be obtained from Brauer diagram considerations.

Bowman, Enyang, and Goodman [8] adopt a cellular basis approach to describing the kernels in the orthogonal and symplectic cases, as well as in the case of the general linear group GL_n acting on mixed tensor powers $\mathsf{V}^{\otimes k} \otimes (\mathsf{V}^*)^{\otimes \ell}$ of its natural n-dimensional module V and its dual V^*. A surjection of the walled Brauer algebra $\mathsf{B}_{k,\ell}(n) \to \mathsf{End}_{\mathsf{GL}_n}(\mathsf{V}^{\otimes k} \otimes (\mathsf{V}^*)^{\otimes \ell})$ is used for this purpose (The algebra $\mathsf{B}_{k,\ell}(n)$ and some of its representation-theoretic properties including its action on $\mathsf{V}^{\otimes k} \otimes (\mathsf{V}^*)^{\otimes \ell}$ can be found, e.g., in [2].).

Although this article is largely a survey discussing recent work on partition algebras and the fundamental theorems of invariant theory for symmetric groups, it features new results. Among them are a new bijection between vacillating tableaux and set-partition tableaux, a new expression for the characters of partition algebras, and new identities relating Stirling numbers of the second kind and fixed points of permutations.

Our main point of emphasis is that the Schur–Weyl duality afforded by the surjection $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ furnishes an effective framework for studying the symmetric group S_n and its invariant theory. Here n is fixed, and the tensor power k is allowed to grow arbitrarily large. When $2k > n$, the kernel of $\Phi_{k,n}$ can best be

described using the orbit basis of the partition algebra and can be used to give the Second Fundamental Theorem of Invariant Theory for S_n.

2 Restriction–Induction Bratteli Diagrams and Vacillating Tableaux

In this section, we discuss the restriction and induction functors for a group–subgroup pair (G, H) and the Bratteli diagram that comes from applying them and then specialize to the case of the pair $(\mathsf{S}_n, \mathsf{S}_{n-1})$. Further details can be found, for example, in [4, Sect. 4].

2.1 Generalities on Restriction and Induction

Let (G, H) be a pair consisting of a finite group G and a subgroup H of G. Let U^0 be the trivial one-dimensional G-module. For $k \in \mathbb{Z}_{\geq 0}$, construct the H-module $\mathsf{U}^{k+\frac{1}{2}}$ by restricting U^k to H and then construct the G-module U^{k+1} by inducing $\mathsf{U}^{k+\frac{1}{2}}$ to G. Thus,

$$\mathsf{U}^{k+\frac{1}{2}} = \mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{U}^k) \qquad \text{and} \qquad \mathsf{U}^{k+1} = \mathsf{Ind}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{U}^{k+\frac{1}{2}}) = \mathbb{F}\mathsf{G} \otimes_{\mathbb{F}\mathsf{H}} \mathsf{U}^{k+\frac{1}{2}}. \tag{2.1}$$

In this way, U^ℓ is defined inductively for all $\ell \in \frac{1}{2}\mathbb{Z}_{\geq 0}$, and

$$\mathsf{U}^k = \left(\mathsf{Ind}^{\mathsf{G}}_{\mathsf{H}}\mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}\right)^k(\mathsf{U}^0) \qquad \text{for all } k \in \mathbb{Z}_{\geq 0}. \tag{2.2}$$

In particular, the module $\mathsf{M} := \mathsf{Ind}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{U}^0)) = \mathsf{U}^1$ is isomorphic to G/H as a G-module, where G acts on the left cosets of G/H by multiplication.

For a G-module X and an H-module Y, the “tensor identity” says that

$$\mathsf{Ind}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{X}) \otimes \mathsf{Y}) \cong \mathsf{X} \otimes \mathsf{Ind}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{Y}). \tag{2.3}$$

The mapping $(g \otimes_{\mathbb{F}\mathsf{H}} x) \otimes y \mapsto gx \otimes (g \otimes_{\mathbb{F}\mathsf{H}} y)$ can be used to establish this isomorphism (see, e.g., [19, (3.18)]). Hence, when $\mathsf{X} = \mathsf{U}^k$ and $\mathsf{Y} = \mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{U}^0)$, (2.3) implies that

$$\mathsf{Ind}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{U}^k)) \cong \mathsf{Ind}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{U}^k) \otimes \mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{U}^0)) \cong \mathsf{U}^k \otimes \mathsf{Ind}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{U}^0)) = \mathsf{U}^k \otimes \mathsf{M}.$$

Therefore, by induction, we have the following isomorphisms for all $k \in \mathbb{Z}_{\geq 0}$:

$$\mathsf{M}^{\otimes k} \cong \mathsf{U}^k \quad (\text{as } \mathsf{G}\text{-modules}) \qquad \text{and} \qquad \mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{M}^{\otimes k}) \cong \mathsf{U}^{k+\frac{1}{2}} \quad (\text{as } \mathsf{H}\text{-modules}). \tag{2.4}$$

The *centralizer algebra of the* G *-action on* $\mathsf{M}^{\otimes k}$ is defined as

$$\mathsf{Z}_{k,\mathsf{G}} := \mathsf{End}_{\mathsf{G}}(\mathsf{M}^{\otimes k}) = \{\varphi \in \mathsf{End}(\mathsf{M}^{\otimes k}) \mid \varphi(g.x) = g.\varphi(x) \text{ for all } x \in \mathsf{M}^{\otimes k}, g \in \mathsf{G}\},$$

and the centralizer algebra for the restriction to H is denoted by $\mathsf{Z}_{k+\frac{1}{2},\mathsf{H}} := \mathsf{End}_{\mathsf{H}}(\mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{M}^{\otimes k}))$. As a consequence of (2.4), we have algebra isomorphisms

$$\begin{aligned} \mathsf{Z}_{k,\mathsf{G}} &= \mathsf{End}_{\mathsf{G}}(\mathsf{M}^{\otimes k}) \cong \mathsf{End}_{\mathsf{G}}(\mathsf{U}^{k}) \quad \text{and} \\ \mathsf{Z}_{k+\frac{1}{2},\mathsf{H}} &= \mathsf{End}_{\mathsf{H}}(\mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{M}^{\otimes k})) \cong \mathsf{End}_{\mathsf{H}}(\mathsf{U}^{k+\frac{1}{2}}). \end{aligned} \tag{2.5}$$

For $k \in \mathbb{Z}_{\geq 0}$ we adopt the following notation to index the irreducible summands in these modules.

- $\Lambda_{k,\mathsf{G}} \subseteq \Lambda_{\mathsf{G}}$ indexes the irreducible G-modules, and hence also the irreducible $\mathsf{Z}_{k,\mathsf{G}}$-modules, occurring in $\mathsf{U}^k \cong \mathsf{M}^{\otimes k}$;
- $\Lambda_{k+\frac{1}{2},\mathsf{H}} \subseteq \Lambda_{\mathsf{H}}$ indexes the irreducible H-modules, and hence also the irreducible $\mathsf{Z}_{k+\frac{1}{2},\mathsf{H}}$-modules, occurring in $\mathsf{U}^{k+\frac{1}{2}} \cong \mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{M}^{\otimes k})$.

2.2 *The Restriction–Induction Bratteli Diagram*

The *restriction–induction Bratteli diagram* for the pair (G, H) is an infinite, rooted tree $\mathcal{B}(\mathsf{G}, \mathsf{H})$ whose vertices are organized into rows labeled by half integers ℓ in $\frac{1}{2}\mathbb{Z}_{\geq 0}$. For $\ell = k \in \mathbb{Z}_{\geq 0}$, the vertices on row k are the elements of $\Lambda_{k,\mathsf{G}}$, and the vertices on row $\ell = k + \frac{1}{2}$ are the elements of $\Lambda_{k+\frac{1}{2},\mathsf{H}}$. The vertex on row 0 is the root, the label of the trivial G-module, and the vertex on row $\frac{1}{2}$ is the label of the trivial H-module.

The edges of $\mathcal{B}(\mathsf{G}, \mathsf{H})$ are constructed from the restriction and induction rules for $\mathsf{H} \subseteq \mathsf{G}$ as follows. Let $\{\mathsf{G}^\lambda\}_{\lambda \in \Lambda_{\mathsf{G}}}$ and $\{\mathsf{H}^\alpha\}_{\alpha \in \Lambda_{\mathsf{H}}}$ be the sets of irreducible modules for these groups over $\mathbb{F}$. By Frobenius reciprocity, the multiplicity c^λ_α of H^α in the restricted module $\mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{G}^\lambda)$ equals the multiplicity of G^λ in the induced module $\mathsf{Ind}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{H}^\alpha)$, and thus

$$\mathsf{Res}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{G}^\lambda) = \bigoplus_{\alpha \in \Lambda_{\mathsf{H}}} c^\lambda_\alpha\, \mathsf{H}^\alpha \qquad \text{and} \qquad \mathsf{Ind}^{\mathsf{G}}_{\mathsf{H}}(\mathsf{H}^\alpha) = \bigoplus_{\lambda \in \Lambda_{\mathsf{G}}} c^\lambda_\alpha\, \mathsf{G}^\lambda. \tag{2.6}$$

In $\mathcal{B}(\mathsf{G}, \mathsf{H})$, we draw c^λ_α edges from $\lambda \in \Lambda_{k,\mathsf{G}}$ to $\alpha \in \Lambda_{k+\frac{1}{2},\mathsf{H}}$ and from $\alpha \in \Lambda_{k+\frac{1}{2},\mathsf{H}}$ to $\lambda \in \Lambda_{k+1,\mathsf{G}}$. Hence, the Bratteli diagram is constructed in such a way that

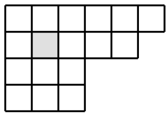
- The number of paths from the root at level 0 to $\lambda \in \Lambda_{k,\mathsf{G}}$ equals the multiplicity of G^λ in $\mathsf{U}^k \cong \mathsf{M}^{\otimes k}$ and thus also equals the dimension of the irreducible $\mathsf{Z}_{k,\mathsf{G}}$-module $\mathsf{Z}^\lambda_{k,\mathsf{G}}$ by (1.5) (these numbers are computed in Pascal-triangle-like fashion and are placed beneath each vertex);

- The number of paths from the root at level 0 to $\alpha \in \Lambda_{k+\frac{1}{2},\mathsf{H}}$ equals the multiplicity of H^α in $\mathsf{U}^{k+\frac{1}{2}}$ and thus also equals the dimension of the $\mathsf{Z}_{k+\frac{1}{2},\mathsf{H}}$-module $\mathsf{Z}^\alpha_{k+\frac{1}{2},\mathsf{H}}$ (and is indicated below each vertex);
- The sum of the squares of the labels on row k (resp. row $k+\frac{1}{2}$) equals $\dim(\mathsf{Z}_{k,\mathsf{G}})$ (resp. $\dim(\mathsf{Z}_{k+\frac{1}{2},\mathsf{H}})$).

The restriction–induction Bratteli diagram for the pair $(\mathsf{S}_5, \mathsf{S}_4)$ is displayed in Fig. 1.

2.3 Restriction and Induction for the Symmetric Group Pair $(\mathsf{S}_n, \mathsf{S}_{n-1})$

Assume $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_n]$ is a partition of n with parts $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$, and identify λ with its Young diagram. For example, we identify the partition $\lambda = [6, 5, 3, 3] \vdash 17$ with its Young diagram as follows,

$\lambda = [6, 5, 3, 3] =$  .

The *hook length* $h(b)$ of a box b in the diagram is 1 plus the number of boxes to the right of b in the same row, plus the number of boxes below b in the same column. Thus, $h(b) = 1 + 3 + 2 = 6$ for the shaded box in the example above. The dimension f^λ of the irreducible S_n-module S_n^λ can be computed by the hook-length formula,

$$f^\lambda = \frac{n!}{\prod_{b \in \lambda} h(b)}, \tag{2.7}$$

where the denominator is the product of the hook lengths as b ranges over the boxes in the Young diagram of λ. This is equal to the number of standard Young tableaux of shape λ, where a *standard Young tableau* T is a filling of the boxes in the Young diagram of λ with the numbers $\{1, 2, \ldots, n\}$ such that the entries increase in every row from left to right and in every column from top to bottom.

The restriction and induction rules for irreducible symmetric group modules S_n^λ are well known (and can be found, e.g., in [21, Thm. 2.43]):

$$\mathsf{Res}^{\mathsf{S}_n}_{\mathsf{S}_{n-1}}(\mathsf{S}_n^\lambda) = \bigoplus_{\mu = \lambda - \square} \mathsf{S}^\mu_{n-1}, \qquad \mathsf{Ind}^{\mathsf{S}_{n+1}}_{\mathsf{S}_n}(\mathsf{S}_n^\lambda) = \bigoplus_{\kappa = \lambda + \square} \mathsf{S}^\kappa_{n+1}, \tag{2.8}$$

where the first sum is over all partitions μ of $n-1$ obtained from λ by removing a box from the end of a row of the diagram of λ, and the second sum is over all partitions κ of $n+1$ obtained by adding a box to the end of a row of λ.

Applying these rules to the trivial one-dimensional S_n-module $\mathsf{S}_n^{[n]}$, we see that

$$\mathsf{Ind}_{\mathsf{S}_{n-1}}^{\mathsf{S}_n}\big(\mathsf{Res}_{\mathsf{S}_{n-1}}^{\mathsf{S}_n}(\mathsf{S}_n^{[n]})\big) = \mathsf{Ind}_{\mathsf{S}_{n-1}}^{\mathsf{S}_n}\big(\mathsf{S}_{n-1}^{[n-1]}\big) = \mathsf{S}_n^{[n]} \oplus \mathsf{S}_n^{[n-1,1]} \cong \mathsf{M}_n.$$

Thus, in the notation of Sect. 2.1, U^1 is the permutation module M_n, and by (2.4),

$$\mathsf{M}_n^{\otimes k} \cong \mathsf{U}^k \ \ (\text{as } \mathsf{S}_n\text{-modules}) \quad \text{and} \quad \mathsf{Res}_{\mathsf{S}_{n-1}}^{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k}) \cong \mathsf{U}^{k+\frac{1}{2}} \ \ (\text{as } \mathsf{S}_{n-1}\text{-modules}). \tag{2.9}$$

Then (2.5) implies that the centralizer algebras are given by

$$\mathsf{Z}_{k,n} := \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k}) \cong \mathsf{End}_{\mathsf{S}_n}(\mathsf{U}^k), \tag{2.10}$$

$$\mathsf{Z}_{k+\frac{1}{2},n} := \mathsf{End}_{\mathsf{S}_{n-1}}(\mathsf{M}_n^{\otimes k}) \cong \mathsf{End}_{\mathsf{S}_{n-1}}(\mathsf{U}^k), \tag{2.11}$$

where we are writing $\mathsf{Z}_{k,n}$ rather than $\mathsf{Z}_{k,\mathsf{S}_n}$ and $\mathsf{Z}_{k+\frac{1}{2},n}$ instead of $\mathsf{Z}_{k+\frac{1}{2},\mathsf{S}_{n-1}}$ to simplify the notation. (*The use of n in place of the more natural choice of $n-1$ for the second one should be especially noted.*)

From our discussions in Sects. 2.2 and 2.3, we know that the following holds.

If $k, n \in \mathbb{Z}_{\geq 0}$ and $n \geq 1$, then for all $\lambda \in \Lambda_{k,\mathsf{S}_n}$,

$$\dim(\mathsf{Z}_{k,n}^{\lambda}) = \mathsf{m}_{k,n}^{\lambda} = \big|\big\{ \text{ paths in } \mathcal{B}(\mathsf{S}_n, \mathsf{S}_{n-1}) \text{ from } [n] \text{ at level } 0 \text{ to } \lambda \text{ at level } k \big\}\big|. \tag{2.12}$$

In Fig. 1, we display the first several rows of the restriction–induction Bratteli diagram $\mathcal{B}(\mathsf{S}_5, \mathsf{S}_4)$. Below each partition, we record the number of paths from the top of the diagram to that partition. For integer values of k, the number beneath the partition $\lambda \vdash n$ represents the multiplicity $\mathsf{m}_{k,n}^{\lambda}$ of the irreducible S_n-module S_n^{λ} in $\mathsf{M}_n^{\otimes k}$ (with $n = 5$ in this particular example). For values $k + \frac{1}{2}$, it indicates the multiplicity of S_{n-1}^{μ}, $\mu \vdash n-1$, in the restriction of $\mathsf{M}_n^{\otimes k}$ to S_{n-1}. The number on the right of each line is $\dim(\mathsf{Z}_{k,n})$ (or $\dim(\mathsf{Z}_{k+\frac{1}{2},n})$), which is the sum of the squares of the subscripts on the line.

The module M_n is faithful, so by Burnside's theorem every irreducible S_n-module appears in $\mathsf{M}_n^{\otimes k}$ for some $k \geq 0$. In the example of Fig. 1, all of the irreducible S_5-modules appear in $\mathsf{M}_5^{\otimes 4}$ (i.e., all of the partitions of 5 appear in $\mathcal{B}(\mathsf{S}_5, \mathsf{S}_4)$ at level $\ell = 4$). More generally, we can use the Bratteli diagram to establish the following result.

Proposition 2.13 (i) *The irreducible module S_n^{λ}, $\lambda \vdash n$, appears as a summand of $\mathsf{M}_n^{\otimes k}$ if and only if $k \geq |\lambda^{\#}|$ (where $\lambda^{\#}$ is the partition obtained by removing the first part of λ).*

(ii) *All irreducible S_n modules appear in $\mathsf{M}_n^{\otimes k}$ if and only if $k \geq n-1$. Thus,*

$$\Lambda_{k,\mathsf{S}_n} = \Lambda_{\mathsf{S}_n} = \{\lambda \mid \lambda \vdash n\} \quad \textit{if and only if} \quad k \geq n-1. \tag{2.14}$$

Proof (i) Let λ be a partition of n. Successively remove the last box from the bottom row and place it at the end of the first row until the partition has only one row.

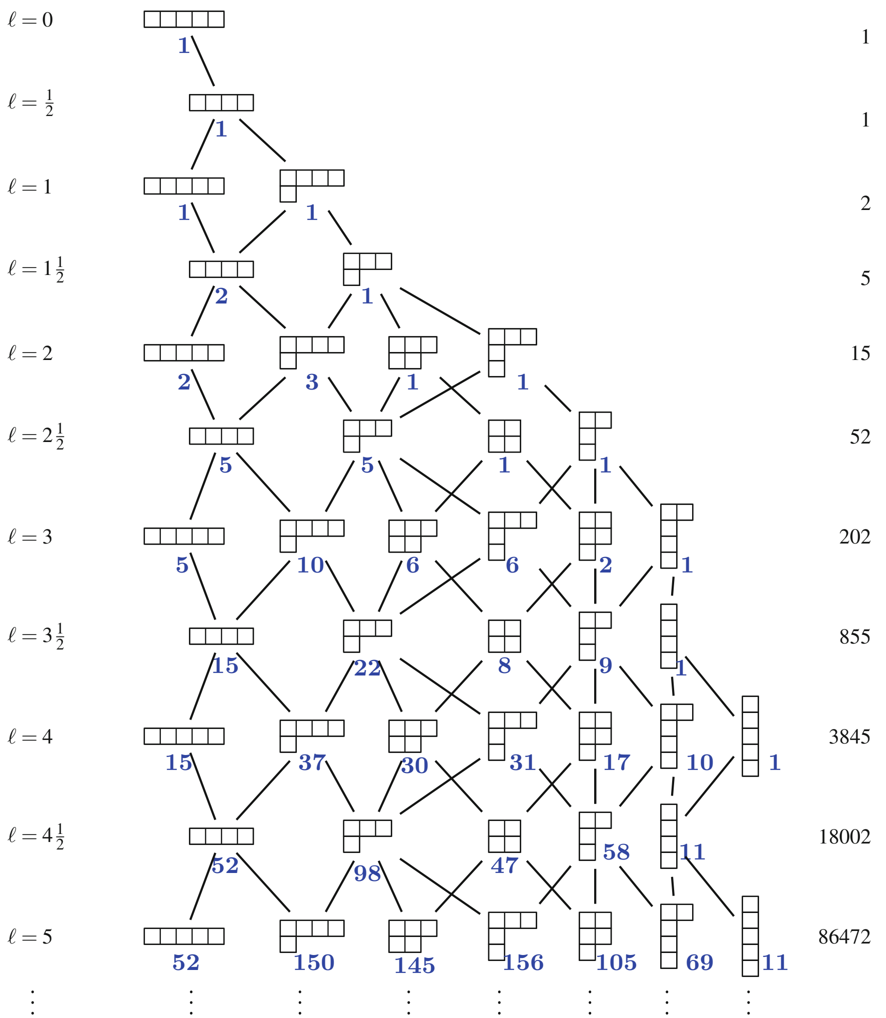
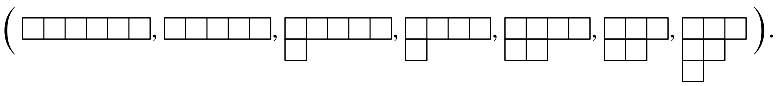


Fig. 1 Levels $0, \frac{1}{2}, 1, \ldots, 3\frac{1}{2}, 4$ of the Bratteli diagram for the pair $(\mathsf{S}_5, \mathsf{S}_4)$. All of the partitions of 5 appear on level $\ell = 4$, and the structure of the diagram stabilizes for $\ell \geq 4$

Counting the removal of a box as one step and the adjoining of the box to the first row as another, this sequence of partitions, when read in reverse order, determines a path in $\mathcal{B}(\mathsf{S}_n, \mathsf{S}_{n-1})$ from $[n]$ at level 0 to λ at level $|\lambda^\#|$. Thus, S_n^λ has multiplicity at least one in $\mathsf{M}_n^{\otimes|\lambda^\#|}$. For example, the path to $\lambda = [3, 2, 1]$ at level $|\lambda^\#| = |[2, 1]| = 3$ in $\mathcal{B}(\mathsf{S}_6, \mathsf{S}_5)$ given by this construction is



It is impossible for λ to appear before level $|\lambda^{\#}|$, since any path from $[n]$ to λ requires removing $|\lambda^{\#}|$ boxes from the first row of the partition and placing them in lower rows. Finally, if $\lambda \in \Lambda_{k,\mathsf{S}_n}$ for some integer $k \geq 0$, then $\lambda \in \Lambda_{k+1,\mathsf{S}_n}$, because we can always remove any removable box in λ and then place it back in its same position. Therefore, S_n^λ occurs in $\mathsf{M}_n^{\otimes k}$ if and only if $k \geq |\lambda^{\#}|$.

(ii) As k increases, the last partition of n to appear in $\mathsf{M}_n^{\otimes k}$ is $\lambda = [1^n]$, which by part (i) first occurs when $k = |[1^n]^{\#}| = n - 1$. Thus, $k \geq n - 1$ is a necessary and sufficient condition for Λ_{k,S_n} to equal Λ_{S_n}. □

Paths in the Bratteli diagram $\mathcal{B}(\mathsf{S}_n, \mathsf{S}_{n-1})$ give the dimension $\mathsf{m}_{k,n}^\lambda$ of the irreducible module $\mathsf{Z}_{k,n}^\lambda$ of the centralizer algebra $\mathsf{Z}_{k,n}$ in the same way that paths in Young's lattice [35, Sect. 5.1] give the dimension of the symmetric group module S_n^λ. For this reason, we make the following definition (see also [4, 10, 18], where this definition is used).

Definition 2.1 Let $k \in \mathbb{Z}_{\geq 0}$ and let $\lambda \in \Lambda_{k,\mathsf{S}_n}$. Each path of length k from $[n]$ at level 0 to λ at level k in the Bratteli diagram $\mathcal{B}(\mathsf{S}_n, \mathsf{S}_{n-1})$ determines a *vacillating tableau* $\underline{\upsilon}$ *of shape* λ *and length* k, which is an alternating sequence

$$\underline{\upsilon} = \left(\lambda^{(0)} = [n],\ \lambda^{(\frac{1}{2})} = [n-1],\ \lambda^{(1)}, \lambda^{(1+\frac{1}{2})}, \ldots, \lambda^{(k-\frac{1}{2})}, \lambda^{(k)} = \lambda\right)$$

of partitions starting at $\lambda^{(0)} = [n]$ and terminating at the partition $\lambda^{(k)} = \lambda$ such that $\lambda^{(i)} \in \Lambda_{i,\mathsf{S}_n}$, $\lambda^{(i+\frac{1}{2})} \in \Lambda_{i+\frac{1}{2},\mathsf{S}_{n-1}}$ for each integer $0 \leq i < k$, and

(a) $\lambda^{(i+\frac{1}{2})} = \lambda^{(i)} - \square$,
(b) $\lambda^{(i+1)} = \lambda^{(i+\frac{1}{2})} + \square$.

In Sect. 3.3, we describe a bijection between vacillating tableaux of shape λ and length k and set-partition tableaux of shape λ whose nonzero entries are $1, 2, \ldots, k$. This bijection is analogous to the well-known bijection between paths to λ of length k in Young's lattice and standard tableaux of shape $\lambda \vdash k$ (see [35, Sect. 5.1]). Since paths in the Bratteli diagram correspond to vacillating tableaux, (2.12) tells us the following:

If $k, n \in \mathbb{Z}_{\geq 0}$ *and* $n \geq 1$, *then for all* $\lambda \in \Lambda_{k,\mathsf{S}_n}$,

$$\dim(\mathsf{Z}_{k,n}^\lambda) = \mathsf{m}_{k,n}^\lambda = \left|\left\{\text{vacillating tableaux of shape } \lambda \text{ and length } k\right\}\right|. \tag{2.15}$$

3 Set Partitions

For $k \in \mathbb{Z}_{\geq 1}$, let $[1, 2k] = \{1, 2, \ldots, 2k\}$, as before. The set partitions in

$$\begin{aligned} \Pi_{2k} &= \{\text{set partitions of } [1, 2k]\}, \\ \Pi_{2k-1} &= \{\pi \in \Pi_{2k} \mid k \text{ and } 2k \text{ are in the same block of } \pi\}. \end{aligned} \tag{3.1}$$

index bases of the partition algebras $\mathsf{P}_k(n)$ and $\mathsf{P}_{k+\frac{1}{2}}(n)$, respectively. We let $|\pi|$ equal the number of blocks of π. For example, if

$$\begin{aligned} \pi &= \left\{1, 8, 9, 10 \,|\, 2, 3 \,|\, 4, 7 \,|\, 5, 6, 11, 12, 14 \,|\, 13\right\} \in \Pi_{14}, \\ \rho &= \left\{1, 8, 9, 10 \,|\, 2, 3 \,|\, 4 \,|\, 5, 6, 11, 12 \,|\, 7, 13, 14\right\} \in \Pi_{13} \subseteq \Pi_{14}, \end{aligned} \tag{3.2}$$

then $\pi \notin \Pi_{13}$ and $|\pi| = |\rho| = 5$.

For $\ell, n \in \mathsf{Z}_{\geq 1}$, define

$$\Pi_{\ell,n} = \{\pi \in \Pi_\ell \mid \pi \text{ has at most } n \text{ blocks}\}. \tag{3.3}$$

The cardinalities of these sets are Bell numbers: $|\Pi_\ell| = \mathsf{B}(\ell)$ and $|\Pi_{\ell,n}| = \mathsf{B}(\ell, n)$. We refer to $\mathsf{B}(\ell, n) = \sum_{t=1}^{n} \left\{ {\ell \atop t} \right\}$ as an *n-restricted Bell number*.

3.1 Multiplicities from a Permutation Module Perspective

We begin by discussing a second approach to decomposing $\mathsf{M}_n^{\otimes k}$ using permutation modules for S_n and then describe connections with set-partition tableaux.

Let $\{\mathsf{v}_1, \mathsf{v}_2, \ldots, \mathsf{v}_n\}$ be the standard permutation basis of M_n such that $\sigma.\mathsf{v}_i = \mathsf{v}_{\sigma(i)}$ for $\sigma \in \mathsf{S}_n$. Assume π is a set partition of $\{1, 2, \ldots, k\}$ into t blocks, where $1 \leq t \leq n$. Then the vector space

$$\mathsf{M}(\pi) := \text{span}_{\mathbb{C}} \left\{\mathsf{v}_{j_1} \otimes \mathsf{v}_{j_2} \otimes \cdots \otimes \mathsf{v}_{j_k} \mid j_a = j_b \iff a, b \text{ are in the same block of } \pi\right\} \tag{3.4}$$

is an S_n-submodule of $\mathsf{M}_n^{\otimes k}$. To see this, recall that $\sigma \in \mathsf{S}_n$ acts diagonally on simple tensors, $\sigma.(\mathsf{v}_{j_1} \otimes \mathsf{v}_{j_2} \otimes \cdots \otimes \mathsf{v}_{j_k}) = \mathsf{v}_{\sigma(j_1)} \otimes \mathsf{v}_{\sigma(j_2)} \otimes \cdots \otimes \mathsf{v}_{\sigma(j_k)}$, and so it preserves the condition in (3.4). As an example, if $n = 8$ and $k = 12$, then

$$\mathsf{v} = \mathsf{v}_3 \otimes \mathsf{v}_1 \otimes \mathsf{v}_3 \otimes \mathsf{v}_3 \otimes \mathsf{v}_4 \otimes \mathsf{v}_5 \otimes \mathsf{v}_4 \otimes \mathsf{v}_1 \otimes \mathsf{v}_1 \otimes \mathsf{v}_3 \otimes \mathsf{v}_4 \otimes \mathsf{v}_5 \in \mathsf{M}_8^{\otimes 12} \tag{3.5}$$

belongs to $\mathsf{M}(\pi)$ for $\pi = \{1, 3, 4, 10 \mid 2, 8, 9 \mid 5, 7, 11 \mid 6, 12\}$.

A simple tensor $\mathsf{v} = \mathsf{v}_{j_1} \otimes \mathsf{v}_{j_2} \otimes \cdots \otimes \mathsf{v}_{j_k} \in \mathsf{M}(\pi)$ also determines an *ordered set partition* of the subscripts $\{1, 2, \ldots, n\}$ into blocks of size $[n - t, 1^t]$ in which the t distinct subscripts of v are in blocks by themselves, in the order they appear in v, and the $n - t$ unused subscripts are in the first block. In the simple tensor v of (3.5), there are $t = 4$ distinct subscripts (namely, 3, 1, 4, and 5, in that order) and

$n - t = 8 - 4 = 4$ unused subscripts $\{2, 6, 7, 8\}$. Such ordered set partitions can be represented by the rows of a *tabloid* of shape $[n - t, 1^t]$ (see [35, Sect. 1.6, 2.1] for details on tabloids). In this particular example, the corresponding tabloid is

$$\{2, 6, 7, 8 \mid 3 \mid 1 \mid 4 \mid 5\} \longleftrightarrow \begin{array}{l} \overline{\underline{2\ 6\ 7\ 8}} \\ \underline{3} \\ \underline{1} \\ \underline{4} \\ \underline{5} \end{array}.$$

The diagonal action of S_n on the simple tensors in $\mathsf{M}(\pi)$ corresponds exactly to the permutation action of S_n on the ordered set partitions of $\{1, 2, \ldots, n\}$ of shape $[n - t, 1^t]$ or, equivalently, on the tabloids of shape $[n - t, 1^t]$. The span of these tabloids is the well-known *permutation module* $\mathsf{M}^{[n-t,1^t]}$ obtained by inducing the trivial module for the subgroup $\mathsf{S}_{n-t} \times \mathsf{S}_1 \times \cdots \times \mathsf{S}_1$ (with t copies of S_1) to S_n (see, e.g., [35, Sect. 2.1]). Thus,

$$\mathsf{M}(\pi) \cong \mathsf{M}^{[n-t,1^t]}, \quad \text{when } \pi \text{ has } t \text{ blocks.}$$

Note that the sizes of the blocks of π are immaterial to this isomorphism, what matters is the number t of blocks. Also observe that since $\mathsf{M}_n = \mathsf{M}^{[n-1,1]}$, the use of the term "permutation module" for both is consistent.

The number of set partitions π of the tensor positions $\{1, 2, \ldots, k\}$ into $1 \le t \le n$ parts is the Stirling number $\left\{ {k \atop t} \right\}$ of the second kind, so by partitioning the simple tensors this way, we obtain the following decomposition of $\mathsf{M}_n^{\otimes k}$ into permutation modules for S_n:

$$\mathsf{M}_n^{\otimes k} \cong \bigoplus_{t=1}^{n} \left\{ {k \atop t} \right\} \mathsf{M}^{[n-t,1^t]}. \tag{3.6}$$

Note that $\left\{ {k \atop t} \right\} = 0$ whenever $t > k$.

Young's rule (see, e.g., [35, Thm. 2.11.2]) gives the decomposition of the permutation module M^{γ} for $\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_n] \vdash n$ into irreducible S_n-modules S_n^{λ}. It states that the multiplicity of S_n^{λ} in M^{γ} equals the Kostka number $\mathsf{K}_{\lambda,\gamma}$. The *Kostka number* $\mathsf{K}_{\lambda,\gamma}$ counts the number of *semistandard tableaux* T of shape λ and type γ, where such a semistandard tableau T is a filling of the boxes of the Young diagram of λ with the numbers $\{0^{\gamma_1}, 1^{\gamma_2}, \ldots, (n-1)^{\gamma_n}\}$, and the entries of T weakly increase across the rows from left to right and strictly increase down the columns. It follows that

$$\mathsf{M}_n^{\otimes k} \cong \bigoplus_{t=1}^{n} \left\{ {k \atop t} \right\} \mathsf{M}^{[n-t,1^t]} \cong \bigoplus_{t=1}^{n} \left\{ {k \atop t} \right\} \left(\sum_{\lambda \vdash n} \mathsf{K}_{\lambda,[n-t,1^t]} \mathsf{S}_n^{\lambda} \right) \cong \bigoplus_{\lambda \vdash n} \left(\sum_{t=1}^{n} \left\{ {k \atop t} \right\} \mathsf{K}_{\lambda,[n-t,1^t]} \right) \mathsf{S}_n^{\lambda}.$$

In this case, the Kostka number $\mathsf{K}_{\lambda,[n-t,1^t]}$ counts the number of semistandard tableaux of shape λ filled with the entries $\{0^{n-t}, 1, 2, \ldots, t\}$. A semistandard tableau of shape λ whose entries are $n - t$ zeros and the numbers $1, 2, \ldots, t$ must have the

$n-t$ zeros in the first row and have a standard filling of the skew shape $\lambda/[n-t]$ with the numbers $1, 2, \ldots, t$. In particular, if $\lambda = [7, 5, 3]$ and $t = 3$ then one such semistandard tableau is

$$\begin{array}{|c|c|c|c|c|c|c|}\hline 0 & 0 & 0 & 3 & 6 & 8 & 12 \\ \hline 1 & 4 & 5 & 7 & 11 \\ \cline{1-5} 2 & 9 & 10 \\ \cline{1-3}\end{array}\,.$$

The number $f^{\lambda/[n-t]}$ of such fillings is given by the hook formula for skew shapes (see, e.g., [36, Cor. 7.16.3]). Furthermore, $\mathsf{K}_{\lambda,[n-t,1^t]} = 0$ whenever $\lambda_1 < n-t$. That is, $\mathsf{K}_{\lambda,[n-t,1^t]} = 0$ whenever $t < n - \lambda_1 = |\lambda^\#|$, where $\lambda^\#$ is the partition obtained by removing one copy of the largest part of λ.

The next result combines [4, Thm. 5.5] with (1.8)–(1.10). In the statement, $\mathsf{M}_n^{\otimes 0}$ should be interpreted as being $\mathbb{F} \cong \mathsf{S}_n^{[n]}$ (as an S_n-module). The discussion above provides a proof of the final equality in part (a) of the following theorem.

Theorem 3.7 *For $k \in \mathbb{Z}_{\geq 0}$ and $n \in \mathbb{Z}_{\geq 1}$, suppose $\mathsf{M}_n^{\otimes k} = \bigoplus_{\lambda \in \Lambda_{k,\mathsf{S}_n}} \mathsf{m}_{k,n}^\lambda \mathsf{S}_n^\lambda$. Assume $\mathsf{Z}_{k,n}^\lambda$ is the irreducible module indexed by λ for $\mathsf{Z}_{k,n} = \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$, as in (1.4)–(1.6).*

(a) *If $\lambda \in \Lambda_{k,\mathsf{S}_n}$, then*

$$\frac{1}{n!}\sum_{\sigma \in \mathsf{S}_n} \mathsf{F}(\sigma)^k \chi_\lambda(\sigma) = \dim(\mathsf{Z}_{k,n}^\lambda) = \mathsf{m}_{k,n}^\lambda = \sum_{t=|\lambda^\#|}^{n} \left\{ {k \atop t} \right\} f^{\lambda/[n-t]},$$

where $\mathsf{F}(\sigma)$ is the number of fixed points of σ, and $f^{\lambda/[n-t]}$ is the number of standard tableaux of skew shape $\lambda/[n-t]$.

(b) $\dim(\mathsf{Z}_{k,n}) = \dim\left(\mathsf{Z}_{2k,n}^{[n]}\right) = \dfrac{1}{n!}\displaystyle\sum_{\sigma \in \mathsf{S}_n} \mathsf{F}(\sigma)^{2k} = \sum_{t=0}^{n} \left\{ {2k \atop t} \right\} = \mathsf{B}(2k, n).$

Furthermore, $\dim(\mathsf{Z}_{k,n}) = \mathsf{B}(2k, n) = \mathsf{B}(2k)$ *if* $n \geq 2k$.

(c) *If $\mu \in \Lambda_{k+\frac{1}{2},\mathsf{S}_{n-1}}$, then*

$$\dim\left(\mathsf{Z}_{k+\frac{1}{2},n}^\mu\right) = \sum_{t=|\mu^\#|}^{n-1} \left\{ {k+1 \atop t+1} \right\} f^{\mu/[n-1-t]} = \frac{1}{(n-1)!}\sum_{\tau \in \mathsf{S}_{n-1}} \left(\mathsf{F}(\tau)+1\right)^k \chi_\mu(\tau).$$

(d) $\dim(\mathsf{Z}_{k+\frac{1}{2},n}) = \dim\left(\mathsf{Z}_{2k+1,n}^{[n-1]}\right) = \displaystyle\sum_{t=1}^{n} \left\{ {2k+1 \atop t} \right\} = \mathsf{B}(2k+1, n).$

Furthermore, $\dim(\mathsf{Z}_{k+\frac{1}{2},n}) = \mathsf{B}(2k+1, n) = \mathsf{B}(2k+1)$ *if* $n \geq 2k+1$.

Remark 3.8 When $n > k$, the top limit in the summation in part (a) can be taken to be k, as the Stirling numbers $\left\{ {k \atop t} \right\}$ are 0 for $t > k$. When $n \leq k$, the term $[n-t, 1^t]$ for $t = n$ is assumed to be the partition $[1^n]$. In that particular case, $\mathsf{K}_{\lambda,[1^n]} = f^\lambda$, the number of standard tableaux of shape λ, as each entry in the tableau appears exactly

once. When $t = n - 1$, the Kostka number is the same, $\mathsf{K}_{\lambda,[1^n]} = f^\lambda$. The only time that the term $t = 0$ contributes is when $k = 0$. In that case, $\left\{ {0 \atop 0} \right\} = 1$, and the Kostka number $\mathsf{K}_{\lambda,[n]} = 0$ if $\lambda \neq [n]$ and $\mathsf{K}_{[n],[n]} = 1$. Thus, $\dim \mathsf{Z}_0^\lambda(n) = \delta_{\lambda,[n]}$, as expected, since $\mathsf{M}_n^{\otimes 0} = \mathsf{S}_n^{[n]}$.

In items (c) and (d) of Theorem 3.7, we have used the fact that as an S_{n-1}-module, $\mathsf{M}_n \cong \mathsf{M}_{n-1} \oplus \mathbb{F}\mathsf{v}_n$, so the character value $\chi_{\mathsf{M}_n}(\tau)$ of $\tau \in \mathsf{S}_{n-1}$ equals 1 plus the number of fixed points of τ viewed as a permutation of $\{1, \ldots, n-1\}$. Then since $\mathsf{Z}_{k+\frac{1}{2},n} = \mathsf{End}_{\mathsf{S}_{n-1}}(\mathsf{M}_n^{\otimes k})$ and M_n is self-dual as an S_{n-1}-module, the first line of part (d) holds. The expressions in (a) and (c) can be related by the next result.

Proposition 3.9 *For all* $k \in \mathbb{Z}_{\geq 0}$,

$$\begin{aligned} \frac{1}{n\,!} \sum_{\sigma \in \mathsf{S}_n} \mathsf{F}(\sigma)^{k+1} &= \mathsf{m}_{k+1,n}^{[n]} = \dim\left(\mathsf{Z}_{k+1,n}^{[n]}\right) \\ &= \dim\left(\mathsf{Z}_{k+\frac{1}{2},n}^{[n-1]}\right) = \mathsf{m}_{k+\frac{1}{2},n-1}^{[n-1]} = \frac{1}{(n-1)!} \sum_{\tau \in \mathsf{S}_{n-1}} \left(\mathsf{F}(\tau)+1\right)^k, \end{aligned} \tag{3.10}$$

where $\mathsf{m}_{k+1,n}^{[n]}$ *is the multiplicity of* $\mathsf{S}_n^{[n]}$ *in* $\mathsf{M}_n^{\otimes(k+1)}$*, and* $\mathsf{m}_{k+\frac{1}{2},n-1}^{[n-1]}$ *is the multiplicity of* $\mathsf{S}_{n-1}^{[n-1]}$ *in* $\mathsf{M}_n^{\otimes k}$*, viewed as an* S_{n-1}*-module.*

Proof The equalities in the first line of (3.10) are a consequence of taking $\lambda = [n]$ in part (a) of Theorem 3.7, and the equalities in the second line come from setting μ equal to $[n-1]$ in part (c). Now observe that the only way $\mathsf{S}_n^{[n]}$ can be obtained by inducing the S_{n-1}-module $\mathsf{Res}_{\mathsf{S}_{n-1}}^{\mathsf{S}_n}\left(\mathsf{M}_n^{\otimes k}\right)$ to S_n is from a summand $\mathsf{S}_{n-1}^{[n-1]}$, so the multiplicity of $\mathsf{S}_n^{[n]}$ in $\mathsf{M}_n^{\otimes(k+1)} = \mathsf{Ind}_{\mathsf{S}_{n-1}}^{\mathsf{S}_n}\left(\mathsf{Res}_{\mathsf{S}_{n-1}}^{\mathsf{S}_n}\left(\mathsf{M}_n^{\otimes k}\right)\right)$, which equals $\dim\left(\mathsf{Z}_{k+1,n}^{[n]}\right)$, is the same as the multiplicity of $\mathsf{S}_{n-1}^{[n-1]}$ in $\mathsf{Res}_{\mathsf{S}_{n-1}}^{\mathsf{S}_n}\left(\mathsf{M}_n^{\otimes k}\right)$, which equals $\dim\left(\mathsf{Z}_{k+\frac{1}{2},n}^{[n-1]}\right)$. This can also be seen in the Bratteli diagram $\mathcal{B}(\mathsf{S}_n, \mathsf{S}_{n-1})$, as there is a unique path between $[n-1]$ at level $k+\frac{1}{2}$ and $[n]$ at level $k+1$ (compare Fig. 1). $\square$

We know from Theorem 3.7(b) that the next result holds for even values of ℓ. Proposition 3.9 combined with part (d) of the theorem allows us to conclude that it holds for odd values of ℓ as well. Consequently, we have the following:

Corollary 3.11 *For all* $\ell \in \mathbb{Z}_{\geq 0}$ *and* $n \in \mathbb{Z}_{\geq 1}$,

$$\mathsf{B}(\ell, n) = \sum_{t=0}^{n} \left\{ {\ell \atop t} \right\} = \frac{1}{n!} \sum_{\sigma \in \mathsf{S}_n} \mathsf{F}(\sigma)^\ell. \tag{3.12}$$

Remark 3.13 The identity in (3.12) has connections with moments of random permutations. If the random variable X denotes the number of fixed points of a uniformly distributed random permutation of a set of size $\ell \geq 1$ into no more than n parts, then the ℓth moment of X is

$$E(X^\ell) = \sum_{t=1}^{n} \left\{ \begin{matrix} \ell \\ t \end{matrix} \right\} = \mathsf{B}(\ell, n) = \frac{1}{n!} \sum_{\sigma \in \mathsf{S}_n} \mathsf{F}(\sigma)^\ell.$$

This formula can be derived by applying Polya's Theorem [25, I, 2, Ex. 9] or by using partition algebra characters as is done in [14, (18)] for $n \geq 2k$.

Remark 3.14 The results of this section relate Stirling numbers of the second kind to the number of fixed points of permutations. In the case of parts (a) and (c) of Theorem 3.7, the expressions involve Stirling numbers and also the number of standard tableaux of certain skew shapes. It would be interesting to have combinatorial bijections that demonstrate the identities. A determinantal formula for the number $f^{\lambda/\nu}$ of standard tableaux of skew shape λ/ν was given by Aitken (see [36, Cor. 7.16.3]). Hook-length formulas for $f^{\lambda/\nu}$ are studied in [30] and the references therein.

3.2 Set-Partition Tableaux

Part (a) of the Theorem 3.7 has inspired the following definition:

Definition 3.1 For $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_n]$ a partition of n, assume $\lambda^\# = [\lambda_2, \ldots, \lambda_n]$ and $t \in \mathbb{Z}$ is such that $|\lambda^\#| \leq t \leq n$. A *set-partition tableau* T *of shape* λ *and content* $\{0^{n-t}, 1, \ldots, k\}$ is a filling of the boxes of λ so that the following requirements are met:

(i) the first $n - t$ boxes of the first row of λ are filled with 0;
(ii) the boxes of the skew shape $\lambda/[n - t]$ are filled with the numbers in $[1, k] = \{1, 2, \ldots, k\}$ such that the entries in each box of $\lambda/[n - t]$ form a block of a set partition $\pi(\mathsf{T})$ of $[1, k]$ having t blocks;
(iii) the boxes of T in the skew shape $\lambda/[n - t]$ strictly increase across the rows and down the columns of $\lambda/[n - t]$, where if b_1 and b_2 are two boxes of $\lambda/[n - t]$, then $b_1 < b_2$ holds if the maximum entries in these boxes satisfy $\mathsf{max}(b_1) < \mathsf{max}(b_2)$.

Example 3.15 Below is a set-partition tableau T of shape $\lambda = [5, 4, 2, 1] \vdash 12$ and content $\{0^4, 1, 2, \ldots, 20\}$ with corresponding set partition $\pi(\mathsf{T}) = \{1, \underline{6} \mid 4, 7, 9, \underline{10} \mid 2, 11, \underline{12} \mid 8, \underline{14} \mid 15, \underline{16} \mid 5, 13, \underline{18} \mid 3, 17, \underline{19} \mid \underline{20}\} \in \Pi_{20}$ consisting of $t = 8$ blocks. The blocks of $\pi(\mathsf{T})$ are listed in increasing order according to their largest elements, which are the underlined numbers.

$$\mathsf{T} = \begin{array}{|c|c|c|c|c|} \hline 0 & 0 & 0 & 0 & 3, 17, \underline{19} \\ \hline 1, \underline{6} & 4, 7, 9, \underline{10} & 5, 13, \underline{18} & \underline{20} \\ \cline{1-4} 2, 11, \underline{12} & 8, \underline{14} \\ \cline{1-2} 15, \underline{16} \\ \cline{1-1} \end{array}$$

Remark 3.16 If λ is a partition of n with $|\lambda^{\#}| = k$ and T is a set-partition tableau of shape λ and content $\{0^{n-t}, 1, \dots, k\}$, then the k boxes of T corresponding to $\lambda^{\#}$ form a set partition of $[1, k]$, so they must be a standard tableau of shape $\lambda^{\#}$. The first row of T must then consist of a single row of zeros of length of $t = \lambda_1 = n - |\lambda^{\#}| = n - k$. For example, if $n = 9$ and $k = 5$, then one possible set-partition tableau of shape λ with $\lambda^{\#} = [3, 2]$ is

$$\begin{array}{|c|c|c|c|} \hline 0 & 0 & 0 & 0 \\ \hline \underline{1} & \underline{3} & \underline{5} & \multicolumn{1}{c}{} \\ \cline{1-3} \underline{2} & \underline{4} & \multicolumn{2}{c}{} \\ \cline{1-2} \end{array}.$$

The following statement is an immediate consequence of Definition 3.1 and Theorem 3.7(a):

If $k, n \in \mathbb{Z}_{\geq 0}$ and $n \geq 1$, then for all $\lambda \in \Lambda_{k,\mathsf{S}_n}$,

$$\dim(\mathsf{Z}_{k,n}^{\lambda}) = \mathsf{m}_{k,n}^{\lambda} = \left| \left\{ \begin{array}{l} \text{set-partition tableaux of shape } \lambda \text{ and content} \\ \{0^{n-t}, 1, \dots, k\} \text{ for some } t \text{ such that } |\lambda^{\#}| \leq t \leq n \end{array} \right\} \right|. \quad (3.17)$$

3.3 Bijections

Combining the results of the previous two sections with what we know from Schur–Weyl duality (1.4)–(1.7), we have the following:

Theorem 3.18 *For $\mathsf{Z}_{k,n} = \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ and for $\lambda \vdash n$, the following are equal:*

(a) *the multiplicity $\mathsf{m}_{k,n}^{\lambda}$ of S_n^{λ} in $\mathsf{M}_n^{\otimes k}$,*
(b) *the dimension of the irreducible $\mathsf{Z}_{k,n}$-module $\mathsf{Z}_{k,n}^{\lambda}$ indexed by λ,*
(c) *the number of paths in the Bratteli diagram $\mathcal{B}(\mathsf{S}_n, \mathsf{S}_{n-1})$ from $[n]$ at level 0 to λ at level k,*
(d) *the number of vacillating tableaux of shape λ and length k,*
(e) *the number of pairs (π, S), where π is a set partition of $\{1, 2, \dots, k\}$ with t blocks, and S is a standard tableau of shape $\lambda/[n-t]$ for some t such that $|\lambda^{\#}| \leq t \leq n$,*
(f) *the number of set-partition tableaux of shape λ and content $\{0^{n-t}, 1, \dots, k\}$, for some t such that $|\lambda^{\#}| \leq t \leq n$.*

The fact that (c) and (d) have the same cardinality is immediate from the definition of vacillating tableaux. The fact that (e) and (f) have equal cardinalities can be seen by taking a set-partition tableau as in (e), and replacing the entries in the boxes having nonzero entries with the numbers $1, 2, \dots, t$ according to their maximal entries from smallest to largest. The reverse process fills the boxes of the standard tableau S with the entries in the blocks of π according to their maximal elements with 1 for the

block with smallest maximal entry and proceeding to t for the block with the largest entry. For instance, the set-partition tableau T of Example 3.15 corresponds to the pair (π, S) given by

$$\pi = \{1, \underline{6} \mid 4, 7, 9, \underline{10} \mid 2, 11, \underline{12} \mid 8, \underline{14} \mid 15, \underline{16} \mid 5, 13, \underline{18} \mid 3, 17, \underline{19} \mid \underline{20}\}, \quad \text{and}$$

$$\mathsf{S} = \begin{array}{|c|c|c|c|c}\cline{5-5} \multicolumn{4}{c|}{} & 7 \\ \hline 1 & 2 & 6 & 8 & \multicolumn{1}{c}{} \\ \hline 3 & 4 & \multicolumn{3}{c}{} \\ \cline{1-2} 5 & \multicolumn{4}{c}{} \\ \cline{1-1} \end{array}.$$

We now describe an algorithm that gives a bijection between set-partition tableaux of shape λ and content $\{0^{n-t}, 1, \dots, k\}$ for some $|\lambda^{\#}| \le t \le n$ and vacillating tableaux of length k. The algorithm assumes familiarity with Schensted row insertion (see [37, Sect. 7.11]). We use $\mathsf{T} \leftarrow b$ to mean row insertion of the box b (along with its entries) into the set-partition tableau T governed by their maximum elements as in Definition 3.1. That is, do usual Schensted insertion on the maximal elements of each box, but then also include all the entries of the box.

A. Set-partition Tableaux ⇒ Vacillating Tableaux

Given a set-partition tableau T of shape $\lambda \vdash n$ and content $\{0^{n-t}, 1, \dots, k\}$, with $|\lambda^{\#}| \le t \le n$, the following algorithm recursively produces a vacillating k-tableau $([n] = \lambda^{(0)}, \lambda^{(\frac{1}{2})}, \lambda^{(1)}, \dots, \lambda^{(k)} = \lambda)$ of shape λ. Examples can be found in Figs. 2 and 3.

(1) Let $\lambda^{(k)} = \lambda$, and set $\mathsf{T}^{(k)} = \mathsf{T}$.
(2) For $j = k, k-1, \dots, 1$ (in descending order), do the following:
   (a) Let $\mathsf{T}^{(j-\frac{1}{2})}$ be the tableau obtained from $\mathsf{T}^{(j)}$ by removing the box b that contains j. At this stage, j will be the largest entry of T so this box will be removable. Let $\lambda^{(j-\frac{1}{2})}$ be the shape of $\mathsf{T}^{(j-\frac{1}{2})}$.
   (b) Delete the entry j from b. If b is then empty, add 0 to it.
   (c) Let $\mathsf{T}^{(j-1)} = \mathsf{T}^{(j-\frac{1}{2})} \leftarrow b$ be the Schensted row insertion of b into $\mathsf{T}^{(j-\frac{1}{2})}$, and let $\lambda^{(j-1)}$ be the shape of $\mathsf{T}^{(j-1)}$.

We delete the largest number j at the jth step, so by our construction $\mathsf{T}^{(j)}$ is a tableau containing a set partition of $\{1, 2, \dots, j\}$ for each $k \ge j \ge 1$. Furthermore, Schensted insertion keeps the rows weakly increasing and columns strictly increasing at each step. At the conclusion, $\mathsf{T}^{(0)}$ is a semistandard tableau that contains only zeros, and as such it must have shape $\lambda^{(0)} = [n]$. The sequence of underlying shapes $(\lambda^{(0)}, \lambda^{(\frac{1}{2})}, \lambda^{(1)}, \dots, \lambda^{(k)})$, listed in reverse order from the way they are constructed, is obtained by removing and adding a box at alternate steps, so it is a vacillating tableau of shape λ and length k.

B. Vacillating Tableaux ⇒ Set-Partition Tableaux

Algorithm A is easily seen to be invertible. Given a vacillating tableau $(\lambda^{(0)}, \lambda^{(\frac{1}{2})}, \lambda^{(1)}, \dots, \lambda^{(k)})$ of shape λ and length k, the following algorithm produces a set-partition

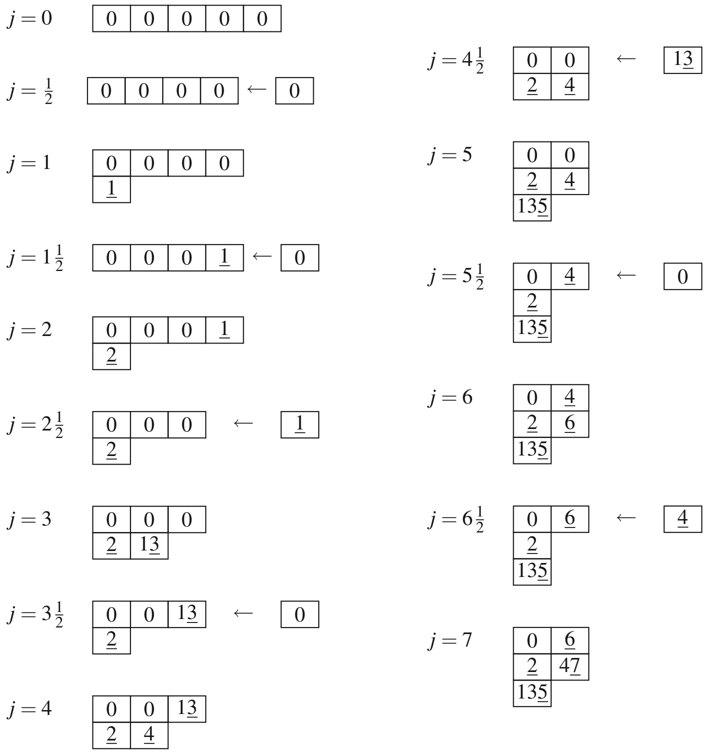


Fig. 2 Bijection between a vacillating 7-tableau of shape [2, 2, 1] and a 7-set-partition tableau of shape [2, 2, 1]

tableau T of shape λ. This process recursively fills the boxes of the shapes $\lambda^{(j)}$ to produce the same fillings as the algorithm above.

(1′) Let $\mathsf{T}^{(0)}$ be the semistandard tableau of shape $\lambda^{(0)} = [n]$ with each of its boxes filled with 0.

(2′) For $j = 0, 1, \ldots, k$, do the following:

  (a′) Let $\mathsf{T}^{(j+\frac{1}{2})}$ be the tableau given by un-inserting the box of $\mathsf{T}^{(j)}$ at position $\lambda^{(j)}/\lambda^{(j+\frac{1}{2})}$, and let b be the box that is un-inserted in this process. That is, $\mathsf{T}^{(j+\frac{1}{2})}$ and b are the unique tableau of shape $\lambda^{(j+\frac{1}{2})}$ and box, respectively, such that $\mathsf{T}^{(j)} = \mathsf{T}^{(j+\frac{1}{2})} \leftarrow b$.

  (b′) Add j to box b. If b contains 0, delete 0 from it.

  (c′) Add the content of the box b to the box in position $\lambda^{(j+1)}/\lambda^{(j+\frac{1}{2})}$, and fill the rest of $\mathsf{T}^{(j+1)}$ with the same entries as in the boxes of $\mathsf{T}^{(j+\frac{1}{2})}$.

Algorithms A and B invert one another step-by-step, since (a) and (c′), (b) and (b′), and (c) and (a′) are easily seen to be the inverses of one another. The fact that steps (a) and (c′) are inverses comes from the fact that Schensted insertion is invertible.

Our bijection implies the following result:

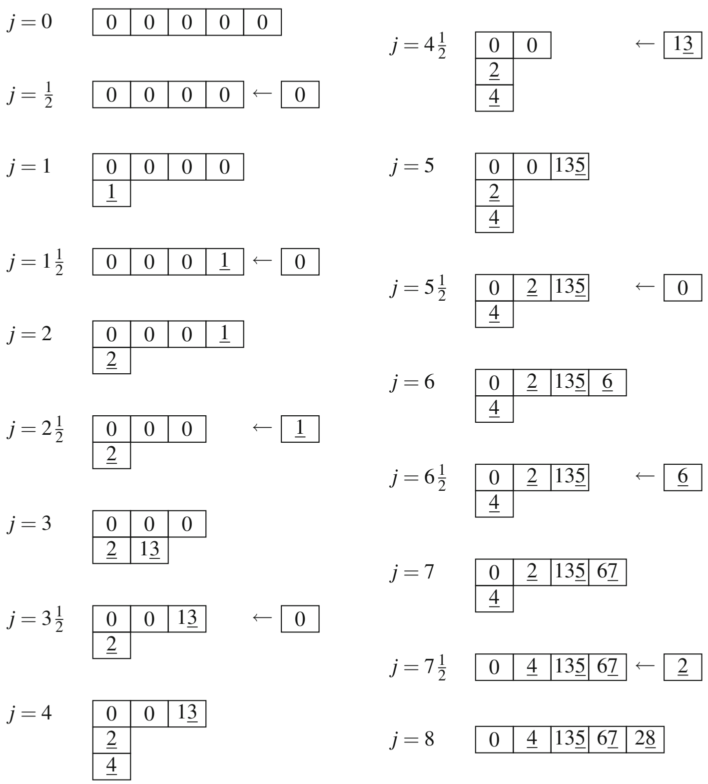


Fig. 3 Bijection between the set partition $\pi = \{4 \mid 135 \mid 67 \mid 28\}$ and an 8-vacillating tableau of shape [5] illustrating Corollary 3.20

Theorem 3.19 *For each $\lambda \in \Lambda_{k,\mathsf{S}_n}$, there is a bijection between the set of vacillating tableaux of shape λ and length k and the set of set-partition tableaux of shape λ and content $\{0^{n-t}, 1, \ldots, k\}$ for some t such that $|\lambda^{\#}| \leq t \leq n$.*

Our bijection also provides a combinatorial proof of the dimension formula for $\mathsf{Z}_{k,n}$ (see (b) of Theorem 3.7), which is illustrated with an example in Fig. 3. In this case, the bijection is between set partitions of $[1, 2k]$ with at most n blocks and vacillating tableaux of shape $[n]$ and length $2k$.

Corollary 3.20 *For $k, n \in \mathbb{Z}_{\geq 1}$, $\dim(\mathsf{Z}_{k,n}) = \dim(\mathsf{Z}^{[n]}_{2k,n}) = \mathsf{m}^{[n]}_{2k,n} = \mathsf{B}(2k, n) = \sum_{t=1}^{n} \left\{ {2k \atop t} \right\}$, the number of set partitions of a set of size $2k$ into at most n parts.*

Proof The first equality comes from (1.7) and the second equality from the previous corollary (or alternatively from (1.6)). To see the third equality, we note that $\mathsf{m}^{[n]}_{2k,n}$ equals the number of set-partition tableaux of shape $[n]$ having content given by a set partition $\pi \in \Pi_{2k}$ with t blocks for $t = 1, 2, \ldots, n$ (the case $t = |[n]^{\#}| = 0$ not

allowed). This is the number of set partitions of $\{1, 2, \ldots, 2k\}$ which have at most n blocks, which is exactly $\mathsf{B}(2k, n)$. $\square$

Remark 3.21 We know that $\dim(\mathsf{Z}_{k,n}) = \dim(\mathsf{Z}_{2k,n}^{[n]}) = \mathsf{m}_{2k,n}^{[n]} = \sum_{\lambda \in \Lambda_{k,\mathsf{S}_n}} (\mathsf{m}_{k,n}^{\lambda})^2$. We can see the last equality combinatorially as well. Given a vacillating tableau $([n], [n-1], \lambda^{(1)}, \ldots, \lambda^{(2k-1)}, [n-1], [n])$ of shape $[n]$ and length $2k$, let $\lambda = \lambda^{(k)}$. Then the first portion of this tableau $([n], [n-1], \lambda^{(1)}, \ldots, \lambda^{(k-\frac{1}{2})}, \lambda^{(k)})$ and the second portion $([n], [n-1], \lambda^{(2k-1)}, \ldots, \lambda^{(k-\frac{1}{2})}, \lambda^{(k)})$ arranged in reverse order form a pair of vacillating tableaux of shape λ and length k. This gives a bijection between vacillating tableaux of shape $[n]$ and length $2k$ and pairs of vacillating tableaux of shape λ and length k for some $\lambda \in \Lambda_{k,\mathsf{S}_n}$.

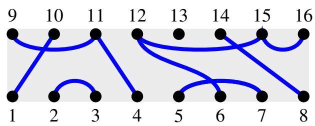
Remark 3.22 Different bijections between set partitions of $[1, 2k]$ and vacillating tableaux of shape $[n]$ and length $2k$ are given in [10, 18]. However, the bijections in those papers require that $n \geq 2k$ holds. The bijection here has the advantage of working for all $k, n \in \mathbb{Z}_{\geq 1}$.

4 The Partition Algebra $\mathsf{P}_k(n)$

4.1 The Diagram Basis of the Partition Algebra $\mathsf{P}_k(n)$

Let $\pi \in \Pi_{2k}$ be a set partition of $[1, 2k] = \{1, 2, \ldots, 2k\}$. The diagram d_π of π has two rows of k vertices each, with the bottom vertices indexed by $1, 2, \ldots, k$, and the top vertices indexed by $k+1, k+2, \ldots, 2k$ from left to right. Edges are drawn so that the connected components of d_π are the blocks of π. An example of a set partition $\pi \in \Pi_{16}$ and its corresponding diagram d_π is

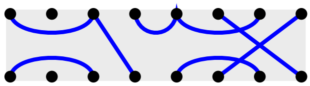
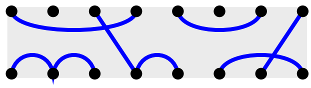
$$\pi = \big\{1, 10 \mid 2, 3 \mid 4, 9, 11 \mid 5, 7 \mid 6, 12, 15, 16 \mid 8, 14 \mid 13\big\} \quad \text{and}$$

$d_\pi =$ 

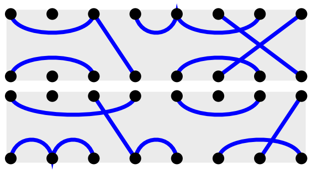
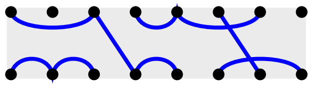
.

The way the edges are drawn is immaterial; what matters is that the connected components of the diagram d_π correspond to the blocks of the set partition π. Thus, d_π represents the equivalence class of all diagrams with connected components equal to the blocks of π.

Multiplication of two diagrams d_{π_1}, d_{π_2} is accomplished by placing d_{π_1} above d_{π_2}, identifying the vertices in the bottom row of d_{π_1} with those in the top row of d_{π_2}, concatenating the edges, deleting all connected components that lie entirely in the middle row of the joined diagrams, and multiplying by a factor of n for each such middle-row component. For example, if

$d_{\pi_1} =$  and $d_{\pi_2} =$ 

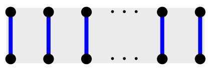
then

$d_{\pi_1} d_{\pi_2} =$  $= n^2$  $= n^2 d_{\pi_1 * \pi_2}$,

where $\pi_1 * \pi_2$ is the set partition obtained by the concatenation of π_1 and π_2 in this process. It is easy to confirm that the product depends only on the underlying set partition and is independent of the diagram chosen to represent π. For any two set partitions $\pi_1, \pi_2 \in \Pi_{2k}$, we let $\langle \pi_1 * \pi_2 \rangle$ denote the number of blocks deleted from the middle of the product $d_{\pi_1} d_{\pi_2}$, so that the product is given by

$$d_{\pi_1} d_{\pi_2} = n^{\langle \pi_1 * \pi_2 \rangle} d_{\pi_1 * \pi_2}. \tag{4.1}$$

For $n \in \mathbb{Z}_{\geq 1}$ and for $k \in \mathbb{Z}_{\geq 1}$, the partition algebra $\mathsf{P}_k(n)$ is the $\mathbb{F}$-span of $\{d_\pi \mid \pi \in \Pi_{2k}\}$ under the diagram multiplication in (4.1). Thus, $\dim(\mathsf{P}_k(n)) = \mathsf{B}(2k)$, the $2k$-th Bell number. We refer to $\{d_\pi \mid \pi \in \Pi_{2k}\}$ as the *diagram basis*. Diagram multiplication is easily seen to be associative with identity element

$$\mathsf{I}_k = \tag{4.2}$$


corresponding to the set partition $\{1, k+1 \mid 2, k+2 \mid \cdots \mid k, 2k\}$.

Set partitions in Π_{2k-1} have k and $2k$ in the same block, and if $\pi_1, \pi_2 \in \Pi_{2k-1}$, then k and $2k$ are also in the same block of $\pi_1 * \pi_2$. Thus, for $k \in \mathbb{Z}_{\geq 1}$, we define $\mathsf{P}_{k-\frac{1}{2}}(n) \subset \mathsf{P}_k(n)$ to be the $\mathbb{F}$-span of $\{d_\pi \mid \pi \in \Pi_{2k-1} \subset \Pi_{2k}\}$. There is also an embedding $\mathsf{P}_k(n) \subset \mathsf{P}_{k+\frac{1}{2}}(n)$ given by adding a top and a bottom node to the right of any diagram in $\mathsf{P}_k(n)$ and a vertical edge connecting them. Setting $\mathsf{P}_0(n) = \mathbb{F}$, we have a tower of embeddings

$$\mathsf{P}_0(n) \cong \mathsf{P}_{\frac{1}{2}}(n) \subset \mathsf{P}_1(n) \subset \mathsf{P}_{1\frac{1}{2}}(n) \subset \mathsf{P}_2(n) \subset \mathsf{P}_{2\frac{1}{2}}(n) \subset \cdots \tag{4.3}$$

with $\dim(\mathsf{P}_k(n)) = |\Pi_{2k}| = \mathsf{B}(2k)$ (the $2k$-th Bell number) for each $k \in \frac{1}{2}\mathbb{Z}_{\geq 1}$.

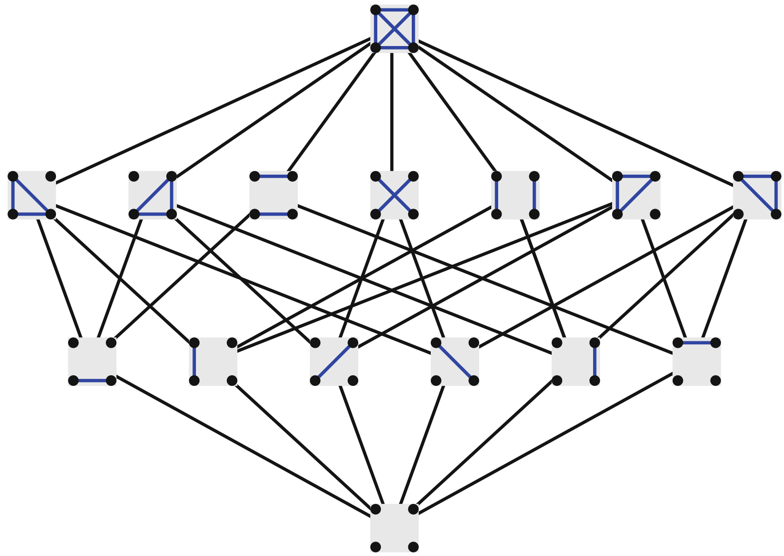


Fig. 4 Hasse diagram of the partition lattice Π_4 in the refinement ordering

4.2 The Orbit Basis

For $k \in \mathbb{Z}_{\geq 1}$, the set partitions Π_{2k} of $[1, 2k]$ form a lattice (a partially ordered set (poset) for which each pair has a least upper bound and a greatest lower bound) under the partial order given by

$$\pi \preceq \rho \quad \text{if every block of } \pi \text{ is contained in a block of } \rho. \tag{4.4}$$

In this case, we say that π is a *refinement* of ρ and that ρ is a *coarsening* of π, so that Π_{2k} is partially ordered by refinement. For example, the Hasse diagram of the partial order $\preceq$ on Π_4 is shown in Fig. 4.

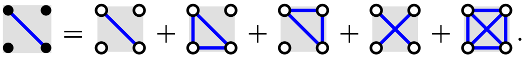
For each $k \in \frac{1}{2}\mathbb{Z}_{\geq 1}$, there is a second basis $\{x_\pi \mid \pi \in \Pi_{2k}\}$ of $\mathsf{P}_k(n)$, called the *orbit basis*, defined by the following coarsening relation with respect to the diagram basis:

$$d_\pi = \sum_{\pi \preceq \rho} x_\rho. \tag{4.5}$$

Thus, the diagram basis element d_π is the sum of all orbit basis elements x_ρ for which ρ is coarser than π. For the remainder of the paper, we adopt the following convention used in [3]:

Diagrams with white vertices indicate orbit basis elements, and those with black vertices indicate diagram basis elements. (4.6)

For example, the expression below writes the diagram $d_{1|23|4}$ in $\mathsf{P}_2(n)$ in terms of the orbit basis,



Remark 4.7 We refer to the basis $\{x_\pi \mid \pi \in \Pi_{2k}\}$ as the orbit basis, because the elements in this basis act on the tensor space $\mathsf{M}_n^{\otimes k}$ in a natural way that corresponds to S_n-orbits on simple tensors (see (5.12)). Jones' original definition of the partition algebra in [22] introduced the orbit basis first and defined the diagram basis later using the refinement relation (4.5). The multiplication rule is more easily stated in the diagram basis, and for that reason the diagram basis is usually the preferred basis when working with $\mathsf{P}_k(n)$. However, when working with $\mathsf{P}_k(n)$ for $2k > n$ the orbit basis is especially useful. For example, in Sect. 5 we are able to describe the kernel of the action of $\mathsf{P}_k(n)$ on tensor space as the two-sided ideal generated by a single orbit basis element.

4.3 *Change of Basis*

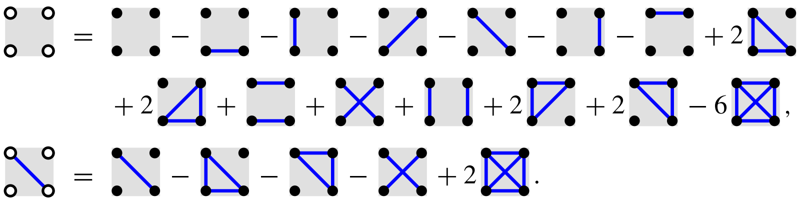
The transition matrix determined by (4.5) between the diagram basis and the orbit basis is the matrix ζ_{2k}, called the *zeta matrix* of the poset Π_{2k}. It is unitriangular with respect to any extension to a linear order, and thus it is invertible, confirming that indeed the elements $x_\pi, \pi \in \Pi_{2k}$, form a basis of $\mathsf{P}_k(n)$. The inverse of ζ_{2k} is the matrix μ_{2k} of the Möbius function of the set-partition lattice, and it satisfies

$$x_\pi = \sum_{\pi \preceq \rho} \mu_{2k}(\pi, \rho) d_\rho, \tag{4.8}$$

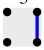
where $\mu_{2k}(\pi, \rho)$ is the (π, ρ) entry of μ_{2k}. The Möbius function of the set-partition lattice can be readily computed using the following formula. If $\pi \preceq \rho$, and ρ consists of ℓ blocks such that the ith block of ρ is the union of b_i blocks of π, then (see, e.g., [36, p. 30]),

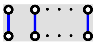
$$\mu_{2k}(\pi, \rho) = \prod_{i=1}^{\ell} (-1)^{b_i - 1}(b_i - 1)!. \tag{4.9}$$

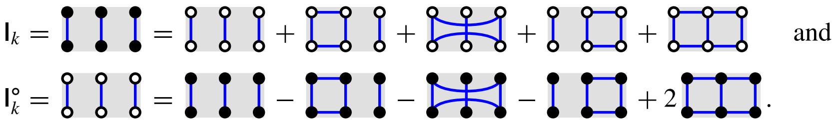
The Hasse diagram of the partition lattice of Π_4 is shown in Fig. 4. In the change of basis between the orbit basis and the diagram basis (in either direction), each basis element is an integer linear combination of the basis elements above or equal to it in the Hasse diagram. If, for example, we apply formula (4.8) to express the orbit basis elements $x_{1\,|\,2\,|\,3\,|\,4}$ and $x_{1\,|\,23\,|\,4}$ in terms of the diagram basis in $\mathsf{P}_2(n)$, we get



(4.10)

Remark 4.11 If $\pi \in \Pi_{2k-1} \subset \Pi_{2k}$, then k and $2k$ are in the same block of π. Since the expression for x_π in the diagram basis and the expression for d_π in the orbit basis are sums over coarsenings of π, they involve only set partitions in Π_{2k-1}. Thus, the relations in (4.5) and (4.8) apply equally well to the algebras $\mathsf{P}_{k-\frac{1}{2}}(n)$. The corresponding Hasse diagram is the sublattice of the diagram for Π_{2k} of partitions greater than or equal to $\{1 \,|\, 2 \,|\, \cdots \,|\, k-1 \,|\, k+1 \,|\, k+2 \,|\, \cdots \,|\, 2k-1 \,|\, k, 2k\}$. For example, the Hasse diagram for Π_3 is found inside that of Π_4 in Fig. 4 as those partitions greater than or equal to .

Remark 4.12 The orbit diagram $\mathsf{I}_k^\circ :=$  is *not* the identity element in $\mathsf{P}_k(n)$. To get the identity element I_k, we must sum all coarsenings of I_k°, as in the first line below:



4.4 Multiplication in the Orbit Basis

In this section, we describe the rule for multiplying in the orbit basis of the partition algebra $\mathsf{P}_k(n)$ using the following conventions:

For $\ell, m \in \mathbb{Z}_{\geq 0}$, let

$$(m)_\ell = m(m-1)\cdots(m-\ell+1). \tag{4.13}$$

Thus, $(m)_\ell = 0$ if $\ell > m$, $(m)_0 = 1$, and $(m)_\ell = m!/(m-\ell)!$ if $m \geq \ell$.

When $\pi \in \Pi_{2k}$, then π induces a set partition on the bottom row $\{1, 2, \ldots, k\}$ and a set partition on the top row $\{k+1, k+2, \ldots, 2k\}$. If $\pi_1, \pi_2 \in \Pi_{2k}$, then we say that $\pi_1 * \pi_2$ *exactly matches in the middle* if the set partition that π_1 induces on its bottom row equals the set partition that π_2 induces on the top row modulo k. For example, if $k = 4$ then $\pi_1 = \{1, 4, 5 \,|\, 2, 8 \,|\, 3, 6, 7\}$ induces the set partition $\{1, 4 \,|\, 2 \,|\, 3\}$ on the bottom row of π_1, and $\pi_2 = \{1, 5, 8 \,|\, 2, 6 \,|\, 3 \,|\, 4, 7\}$ induces the set partition $\{5, 8 \,|\, 6 \,|\, 7\} \equiv \{1, 4 \,|\, 2 \,|\, 3\} \bmod 4$ on the top row of π_2. Thus, $\pi_1 * \pi_2$

exactly matches in the middle. This definition is easy to see in terms of the diagrams, as the examples below, particularly Example 4.16, demonstrate.

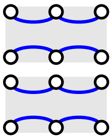
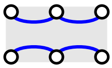
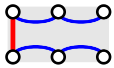
In the next result, we describe the formula for multiplying two orbit basis diagrams. This formula was originally stated by Halverson and Ram in unpublished notes and was proven in [3, Cor. 4.12]. In the product expression below, $x_\rho = 0$ whenever ρ has more than n blocks. Recall that $\langle \pi_1 * \pi_2 \rangle$ is the number of blocks deleted from the middle row of $\pi_1 * \pi_2$.

Theorem 4.14 *Multiplication in* $\mathsf{P}_k(n)$ *in terms of the orbit basis* $\{x_\pi\}_{\pi \in \Pi_{2k}}$ *is given by*

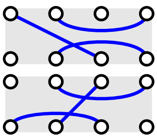
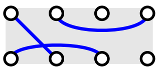
$$x_{\pi_1} x_{\pi_2} = \begin{cases} \displaystyle\sum_\rho (n - |\rho|)_{\langle \pi_1 * \pi_2 \rangle}\, x_\rho, & \text{if } \pi_1 * \pi_2 \text{ exactly matches in the middle,} \\ 0 & \text{otherwise,} \end{cases}$$

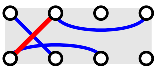
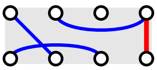
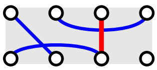
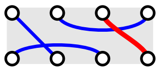
where the sum is over all coarsenings ρ *of* $\pi_1 * \pi_2$ *obtained by connecting blocks that lie entirely in the top row of* π_1 *to blocks that lie entirely in the bottom row of* π_2.

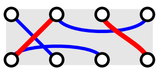
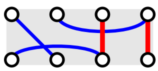
Example 4.15 Suppose $k = 3, n \geq 2$, and $\pi = \{1, 2, 3 \,|\, 4, 5, 6\} \in \Pi_6$. Then, according to Theorem 4.14,

 $= (n-2)$  $+ (n-1)$ , $\quad n \geq 2.$

Example 4.16 Here $k = 4, n \geq 5$, and $\langle \pi_1 * \pi_2 \rangle = 2$ (two blocks are removed upon concatenation of π_1 and π_2).

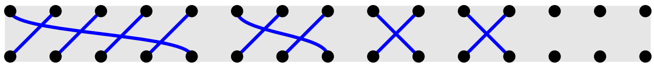
 $= (n-5)(n-6)$ 

$+(n-4)(n-5)\Big($  $+$  $+$  $+$  $\Big)$

$+(n-3)(n-4)\Big($  $+$  $\Big).$

4.5 Characters of Partition Algebras

Let γ_r be the r-cycle $(1, 2, \dots, r)$ in $\mathsf{S}_r \subseteq \mathsf{P}_r(n)$. For a partition $\mu = [\mu_1, \mu_2, \dots, \mu_t] \vdash \ell$, with $\mu_t > 0$ and $0 \le \ell \le k$, define $\gamma_\mu = \gamma_{\mu_1} \otimes \gamma_{\mu_2} \otimes \cdots \otimes \gamma_{\mu_t} \otimes (\substack{\bullet\\ \bullet})^{\otimes(k-\ell)} \in \mathsf{P}_k(n)$, where "$\otimes$" here stands for juxtaposing diagrams from smaller partition algebras. For example, if $\mu = [5, 3, 2, 2]$ and $k = 15$, then the diagram of γ_μ in $\mathsf{P}_k(n)$ is

$\gamma_{[5,3,2,2]} =$  .

In [17, Sect. 2.2], it is shown that characters of the partition algebra $\mathsf{P}_k(n)$ are completely determined by their values on the elements of the set $\{\gamma_\mu \mid \mu \vdash \ell,\ 0 \le \ell \le k\}$, and thus, the γ_μ are analogous to conjugacy class representatives in a group.

For any integer $m \in \mathbb{Z}_{\ge 1}$, let $\mathsf{F}_m(\sigma) = \mathsf{F}(\sigma^m)$, the number of fixed points of σ^m for $\sigma \in \mathsf{S}_n$. For the partition $\mu = [\mu_1, \mu_2, \dots, \mu_t] \vdash \ell$ above, set

$$\mathsf{F}_\mu(\sigma) = \mathsf{F}_{\mu_1}(\sigma) \cdots \mathsf{F}_{\mu_t}(\sigma).$$

Applying the duality between S_n and $\mathsf{P}_k(n)$, Halverson [17, Thm. 3.22] showed that the character value of $\sigma \times \gamma_\mu$ on $\mathsf{M}_n^{\otimes k}$ is $n^{k-\ell}\mathsf{F}_\mu(\sigma)$.

The conjugacy classes of S_n are indexed by the partitions $\delta \vdash n$ that correspond to the cycle type of a permutation. The number of permutations of cycle type δ is $n!/z_\delta$, where $z_\delta = 1^{\delta_1} 2^{\delta_2} \cdots n^{\delta_n}\, \delta_1! \delta_2! \cdots \delta_n!$ when δ has δ_i parts equal to i. The number of fixed points of a permutation depends only on its cycle type. So F_μ is a class function, and we let $\mathsf{F}_\mu(\delta)$ be the value of F_μ on the conjugacy class labeled by δ. Similarly, we let $\chi_\lambda(\delta)$ denote the value of the irreducible S_n-character χ_λ, $\lambda \vdash n$, on the class labeled by δ. Then, applying (1.11), we have the following (compare [17, Cor. 3.25]):

Theorem 4.17 *Assume $n \ge 2k$. For $\lambda \in \Lambda_{k,\mathsf{S}_n}$, and $\mu = [\mu_1, \mu_2, \dots, \mu_t] \vdash \ell$ with $0 \le \ell \le k$, the value of the irreducible character ξ_λ for $\mathsf{P}_k(n)$ on γ_μ is given by*

$$\xi_\lambda(\gamma_\mu) = \frac{n^{k-\ell}}{n!} \sum_{\sigma \in \mathsf{S}_n} \mathsf{F}_\mu(\sigma)\chi_\lambda(\sigma^{-1}) = \frac{n^{k-\ell}}{n!} \sum_{\sigma \in \mathsf{S}_n} \mathsf{F}_\mu(\sigma)\chi_\lambda(\sigma) = n^{k-\ell} \sum_{\delta \vdash n} \frac{1}{z_\delta} \mathsf{F}_\mu(\delta)\chi_\lambda(\delta).$$

Remark 4.18 In the special case that $\mu = [1^k]$ (the partition of k with all parts equal to 1), then $\mathsf{F}_\mu(\sigma) = \mathsf{F}_1(\sigma)^k = \mathsf{F}(\sigma)^k$, and $\gamma_\mu = \mathsf{I}_k$, the identity element of $\mathsf{P}_k(n)$. The result in Theorem 4.17 reduces to

$$\dim \mathsf{Z}_{k,n}^\lambda = \xi_\lambda(\mathsf{I}_k) = \frac{1}{n!} \sum_{\sigma \in \mathsf{S}_n} \mathsf{F}(\sigma)^k \chi_\lambda(\sigma),$$

which is exactly the expression in Theorem 3.7(a) for the dimension of the irreducible $\mathsf{Z}_{k,n}$-module $\mathsf{Z}_{k,n}^{\lambda}$ indexed by λ, since we can identify $\mathsf{P}_k(n)$ with the centralizer algebra $\mathsf{Z}_{k,n} = \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ when $n \geq 2k$.

5 The Representation $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ and Its Kernel

Jones [22] defined an action of the partition algebra $\mathsf{P}_k(n)$ on the tensor space $\mathsf{M}_n^{\otimes k}$ that commutes with the diagonal action of the symmetric group S_n on that same space and showed that this action affords a representation of $\mathsf{P}_k(n)$ onto the centralizer algebra $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$. In this section, we describe the action of each diagram basis element d_π and each orbit basis element x_π of $\mathsf{P}_k(n)$ on $\mathsf{M}_n^{\otimes k}$, and we use the orbit basis to describe the image and the kernel of this action.

5.1 *The Orbit Basis of* $\mathsf{End}_{\mathsf{G}}(\mathsf{M}_n^{\otimes k})$ *for* G *a Subgroup of* S_n

Assume $k, n \in \mathbb{Z}_{\geq 1}$, and let $\{\mathsf{v}_j \mid 1 \leq j \leq n\}$ be the basis for the permutation module M_n of S_n. The elements $\mathsf{v}_{\mathsf{r}} = \mathsf{v}_{r_1} \otimes \cdots \otimes \mathsf{v}_{r_k}$ for $\mathsf{r} = (r_1, \ldots, r_k) \in [1, n]^k = \{1, 2, \ldots, n\}^k$ form a basis for the S_n-module $\mathsf{M}_n^{\otimes k}$ with S_n acting diagonally, $\sigma.\mathsf{v}_{\mathsf{r}} = \mathsf{v}_{\sigma(\mathsf{r})} := \mathsf{v}_{\sigma(r_1)} \otimes \cdots \otimes \mathsf{v}_{\sigma(r_n)}$, as in (1.2).

Suppose $\varphi = \sum_{\mathsf{r},\mathsf{s}\in[1,n]^k} \varphi_{\mathsf{r}}^{\mathsf{s}} \mathsf{E}_{\mathsf{r}}^{\mathsf{s}} \in \mathsf{End}(\mathsf{M}_n^{\otimes k})$ where $\{\mathsf{E}_{\mathsf{r}}^{\mathsf{s}}\}$ is a basis of matrix units for $\mathsf{End}(\mathsf{M}_n^{\otimes k})$, and the coefficients $\varphi_{\mathsf{r}}^{\mathsf{s}}$ belong to $\mathbb{F}$. Then $\mathsf{E}_{\mathsf{r}}^{\mathsf{s}}\mathsf{v}_{\mathsf{t}} = \delta_{\mathsf{r},\mathsf{t}}\mathsf{v}_{\mathsf{s}}$, with $\delta_{\mathsf{r},\mathsf{t}}$ being the Kronecker delta, and the action of φ on the basis of simple tensors is given by

$$\varphi(\mathsf{v}_{\mathsf{r}}) = \sum_{\mathsf{s}} \varphi_{\mathsf{r}}^{\mathsf{s}}\mathsf{v}_{\mathsf{s}}. \tag{5.1}$$

For any subgroup $\mathsf{G} \subseteq \mathsf{S}_n$ (in particular, for S_n itself) and for the centralizer algebra $\mathsf{End}_{\mathsf{G}}(\mathsf{M}_n^{\otimes k}) = \{\varphi \in \mathsf{End}(\mathsf{M}_n^{\otimes k}) \mid \varphi\sigma = \sigma\varphi \text{ for all } \sigma \in \mathsf{G}\}$, we have

$$\begin{aligned}
\varphi \in \mathsf{End}_{\mathsf{G}}(\mathsf{M}_n^{\otimes k}) &\iff \sigma\varphi = \varphi\sigma \quad \text{for all } \ \sigma \in \mathsf{G} \\
&\iff \sum_{\mathsf{s}\in[1,n]^k} \varphi_{\mathsf{r}}^{\mathsf{s}}\, \mathsf{v}_{\sigma(\mathsf{s})} = \sum_{\mathsf{s}\in[1,n]^k} \varphi_{\sigma(\mathsf{r})}^{\mathsf{s}}\, \mathsf{v}_{\mathsf{s}} \quad \text{for all } \ \mathsf{r} \in [1, n]^k,
\end{aligned}$$

and so

$$\varphi \in \mathsf{End}_{\mathsf{G}}(\mathsf{M}_n^{\otimes k}) \iff \varphi_{\mathsf{r}}^{\mathsf{s}} = \varphi_{\sigma(\mathsf{r})}^{\sigma(\mathsf{s})} \quad \text{for all } \ \mathsf{r}, \mathsf{s} \in [1, n]^k, \sigma \in \mathsf{G}. \tag{5.2}$$

It is convenient to view the pair of k-tuples $\mathsf{r}, \mathsf{s} \in [1, n]^k$ in (5.2) as a single $2k$-tuple $(\mathsf{r}, \mathsf{s}) \in [1, n]^{2k}$. In this notation, condition (5.2) tells us that the elements of

$\mathsf{End}_{\mathsf{G}}(\mathsf{M}_n^{\otimes k})$ are in one-to-one correspondence with the G-orbits on $[1, n]^{2k}$, where $\sigma \in \mathsf{G}$ acts on $(r_1, \ldots, r_{2k}) \in [1, n]^{2k}$ by $\sigma(r_1, \ldots, r_{2k}) = (\sigma(r_1), \ldots, \sigma(r_{2k}))$.

We adopt the shorthand notation $(\mathsf{r}|\mathsf{r}') = (r_1, \ldots, r_{2k}) \in [1, n]^{2k}$ when $\mathsf{r} = (r_1, \ldots, r_k) \in [1, n]^k$ and $\mathsf{r}' = (r_{k+1}, \ldots, r_{2k}) \in [1, n]^k$. Let the G-orbit of $(\mathsf{r}|\mathsf{r}') \in [1, n]^{2k}$ be denoted by $\mathsf{G}(\mathsf{r}|\mathsf{r}') = \{\sigma(\mathsf{r}|\mathsf{r}') \mid \sigma \in \mathsf{G}\}$ and define

$$\mathsf{X}_{(\mathsf{r}|\mathsf{r}')} = \sum_{(\mathsf{s}|\mathsf{s}') \in \mathsf{G}(\mathsf{r}|\mathsf{r}')} \mathsf{E}_{\mathsf{s}}^{\mathsf{s}'}, \tag{5.3}$$

where the sum is over the distinct elements in the orbit. This is the indicator function of the orbit $\mathsf{G}(\mathsf{r}|\mathsf{r}')$, and it satisfies (5.2), so $\mathsf{X}_{(\mathsf{r}|\mathsf{r}')} \in \mathsf{End}_{\mathsf{G}}(\mathsf{M}_n^{\otimes k})$. Let $[1, n]^{2k}/\mathsf{G}$ be a set consisting of one $2k$-tuple $(\mathsf{r}|\mathsf{r}')$ for each G-orbit. Since (5.2) is a necessary and sufficient condition for a transformation to belong to $\mathsf{End}_{\mathsf{G}}(\mathsf{M}_n^{\otimes k})$, and the elements $\mathsf{X}_{(\mathsf{r}|\mathsf{r}')}$ for $(\mathsf{r}|\mathsf{r}') \in [1, n]^{2k}/\mathsf{G}$ are linearly independent, we have the following result.

Theorem 5.4 *For* $\mathsf{G} \subseteq \mathsf{S}_n$ *and* $n, k \in \mathbb{Z}_{\geq 1}$, *the centralizer algebra* $\mathsf{End}_{\mathsf{G}}(\mathsf{M}_n^{\otimes k})$ *has a basis* $\{\mathsf{X}_{(\mathsf{r}|\mathsf{r}')} \mid (\mathsf{r}|\mathsf{r}') \in [1, n]^{2k}/\mathsf{G}\}$. *In particular,* $\dim(\mathsf{End}_{\mathsf{G}}(\mathsf{M}_n^{\otimes k}))$ *equals the number of* G*-orbits on* $[1, n]^{2k}$.

When $\mathsf{G} = \mathsf{S}_n$, then since S_n acts transitively on $[1, n]$, the S_n-orbits on $[1, n]^{2k}$ correspond to set partitions of $[1, 2k]$ into *at most n blocks*. In particular, if $\pi \in \Pi_{2k}$ is a set partition of $[1, 2k]$, then

$$\{(r_1, r_2, \ldots, r_{2k}) \mid r_a = r_b \iff a, b \text{ are in the same block of } \pi\} \tag{5.5}$$

is the S_n-orbit corresponding to π (compare this with (3.4)). The condition (5.5) requires there to be n or fewer blocks in π; otherwise there are not enough distinct values $r_a \in [1, n]$ to assign to each of the blocks.

If $\mathsf{G} \subset \mathsf{S}_n$ is a proper subgroup, then the S_n-orbits may split into smaller G-orbits. For example, if $k = 2$ and $\mathsf{G} = \mathsf{A}_4 \subseteq \mathsf{S}_4$, the alternating subgroup, then the S_4-orbit corresponding to $\pi = \{1 \mid 2 \mid 3, 4\}$ contains $(1, 2, 3, 3)$ and $(1, 2, 4, 4)$, but no element of A_4 sends $(1, 2, 3, 3)$ to $(1, 2, 4, 4)$, since it would have to fix 1 and 2 and swap 3 and 4.

When $\mathsf{G} = \mathsf{S}_n$, the orbit basis has an especially nice form, since the S_n-orbits of $[1, n]^{2k}$ correspond to set partitions of $[1, 2k]$ having at most n parts. For $\pi \in \Pi_{2k}$, we designate a special labeling associated with π as follows.

Definition 5.1 Let B_1 be the block of π containing 1, and for $1 < j \leq |\pi|$, let B_j be the block of π containing the smallest number not in $\mathsf{B}_1 \cup \mathsf{B}_2 \cup \cdots \cup \mathsf{B}_{j-1}$. The *standard labeling of* π is $(\mathsf{b}_\pi|\mathsf{b}'_\pi)$, where $\mathsf{b}_\pi = (\mathsf{b}_1, \ldots, \mathsf{b}_k)$ and $\mathsf{b}'_\pi = (\mathsf{b}_{k+1}, \ldots, \mathsf{b}_{2k})$ in $[1, n]^k$, and

$$\mathsf{b}_\ell = j \ \text{ if } \ \ell \in \mathsf{B}_j \ \text{ for } \ \ell \in [1, 2k]. \tag{5.6}$$

Rather than writing $\mathsf{X}_{(\mathsf{b}_\pi|\mathsf{b}'_\pi)}$ for the S_n-orbit element determined by $(\mathsf{b}_\pi|\mathsf{b}'_\pi) \in [1, n]^{2k}$, we denote this simply as X_π. Then it follows that

$$\mathsf{X}_\pi = \sum_{(\mathsf{s}|\mathsf{s}') \in \mathsf{S}_n(\mathsf{b}_\pi|\mathsf{b}'_\pi)} \mathsf{E}_\mathsf{s}^{\mathsf{s}'} = \sum_{(\mathsf{r}|\mathsf{r}')\in[1,n]^{2k}} (\mathsf{X}_\pi)_\mathsf{r}^{\mathsf{r}'} \mathsf{E}_\mathsf{r}^{\mathsf{r}'}, \tag{5.7}$$

where

$$(\mathsf{X}_\pi)_\mathsf{r}^{\mathsf{r}'} = \begin{cases} 1 & \text{if } r_a = r_b \textit{ if and only if } a \text{ and } b \text{ are in the same block of } \pi, \\ 0 & \text{otherwise.} \end{cases} \tag{5.8}$$

In this case, Theorem 5.4 specializes to the following.

Theorem 5.9 *For* $k, n \in \mathbb{Z}_{\geq 1}$, $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ *has a basis* $\{\mathsf{X}_\pi \mid \pi \in \Pi_{2k,n}\}$, *and therefore* $\dim(\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})) = \mathsf{B}(2k, n)$.

Remark 5.10 If G is any finite group and M is any permutation module for G (that is, $g \in \mathsf{G}$ acts as a permutation on a distinguished basis of n elements of M), then G can be viewed as a subgroup of S_n and M can be regarded as the module M_n. Thus, the method of this section applies to tensor powers of any permutation module for any group G. For example, an action of a group G on a finite set $\{x_1, \ldots, x_n\}$ can be viewed as an action of the subgroup G of S_n on the permutation module $\mathrm{span}_\mathbb{F}\{x_1, x_2, \ldots, x_n\}$.

5.2 The Definition of $\Phi_{k,n}$

For $k, n \in \mathbb{Z}_{\geq 1}$, define $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}(\mathsf{M}_n^{\otimes k})$ by

$$\Phi_{k,n}(x_\pi) = \begin{cases} \mathsf{X}_\pi & \text{if } \pi \text{ has } n \text{ or fewer blocks, and} \\ 0 & \text{if } \pi \text{ has more than } n \text{ blocks.} \end{cases} \tag{5.11}$$

As $\{x_\pi \mid \pi \in \Pi_{2k}\}$ is a basis for $\mathsf{P}_k(n)$, we can extend $\Phi_{k,n}$ linearly to get a transformation, $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}(\mathsf{M}_n^{\otimes k})$. It follows from Theorem 5.9 that $\Phi_{k,n}$ maps $\mathsf{P}_k(n)$ surjectively onto $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ for all $k, n \in \mathbb{Z}_{\geq 1}$. When $n \geq 2k$, we have $\dim(\mathsf{P}_k(n)) = \mathsf{B}(2k) = \dim(\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k}))$, and thus $\Phi_{k,n}$ is a bijection.

From (5.11) we see that

$$\Phi_{k,n}(x_\pi)_\mathsf{r}^{\mathsf{r}'} = \begin{cases} 1 & \text{if } r_a = r_b \textit{ if and only if } a \text{ and } b \text{ are in the same block of } \pi, \\ 0 & \text{otherwise.} \end{cases} \tag{5.12}$$

Since the diagram basis $\{d_\pi \mid \pi \in \Pi_{2k}\}$ is related to the orbit basis $\{x_\pi \mid \pi \in \Pi_{2k}\}$ by the refinement relation (4.5), we have as an immediate consequence,

$$\Phi_{k,n}(d_\pi)_\mathsf{r}^{\mathsf{r}'} = \begin{cases} 1 & \text{if } r_a = r_b \textit{ when } a \text{ and } b \text{ are in the same block of } \pi, \\ 0 & \text{otherwise.} \end{cases} \tag{5.13}$$

The map $\Phi_{k,n}$ can be shown to be an algebra homomorphism (the diagram basis works especially well for doing that, see [3, Prop. 3.6]), and so $\Phi_{k,n}$ affords a representation of the partition algebra $\mathsf{P}_k(n)$.

Example 5.14 (Example 4.15 revisited) Recall that in this example $\pi = \{1, 2, 3 \mid 4, 5, 6\}$ and $\rho = \{1, 2, 3, 4, 5, 6 \mid\}$ are elements of Π_6. Then, by definition, $\Phi_{3,n}(x_\pi) = \sum_{i \neq j \in [1,n]} \mathsf{E}_{iii}^{jjj}$ for all $n \geq 2$ so that

$$\begin{aligned}\Phi_{3,n}(x_\pi^2) = \left(\Phi_{3,n}(x_\pi)\right)^2 &= (n-2) \sum_{i \neq j \in [1,n]} \mathsf{E}_{iii}^{jjj} + (n-1) \sum_{i \in [1,n]} \mathsf{E}_{iii}^{iii} \\ &= (n-2)\Phi_{3,n}(x_\pi) + (n-1)\Phi_{3,n}(x_\rho) = \Phi_{3,n}\big((n-2)x_\pi + (n-1)x_\rho\big).\end{aligned}$$

Such examples inspired the product rule in Theorem 4.14.

We identify S_{n-1} with the subgroup of S_n of permutations that fix n and make the identification $\mathsf{M}_n^{\otimes k} \cong \mathsf{M}_n^{\otimes k} \otimes \mathsf{v}_n \subseteq \mathsf{M}_n^{\otimes(k+1)}$, so that $\mathsf{M}_n^{\otimes k}$ is a submodule for both S_{n-1} and $\mathsf{P}_{k+\frac{1}{2}}(n) \subset \mathsf{P}_{k+1}(n)$. Then for tuples $\tilde{\mathsf{r}}, \tilde{\mathsf{s}} \in [1,n]^{k+1}$ having $r_{k+1} = n = s_{k+1}$, condition (5.2) for $\mathsf{G} = \mathsf{S}_{n-1}$ becomes

$$\varphi_{\tilde{\mathsf{r}}}^{\tilde{\mathsf{s}}} = \varphi_{\sigma(\tilde{\mathsf{r}})}^{\sigma(\tilde{\mathsf{s}})} \quad \text{for all } \tilde{\mathsf{r}}, \tilde{\mathsf{s}} \in [1,n]^{k+1}, \sigma \in \mathsf{S}_{n-1}.$$

Thus, the matrix units for $\mathsf{G} = \mathsf{S}_{n-1}$ in (5.3) correspond to set partitions in Π_{2k+1}; that is, set partitions of $\{1, 2, \ldots, 2(k+1)\}$ having $k+1$ and $2(k+1)$ in the same block.

Let $\Phi_{k+\frac{1}{2},n} : \mathsf{P}_{k+\frac{1}{2}}(n) \to \mathsf{End}(\mathsf{M}_n^{\otimes k} \otimes \mathsf{v}_n)$ be defined by

$$\Phi_{k+\frac{1}{2},n}(x_\pi) = \sum_{(\tilde{\mathsf{r}}|\tilde{\mathsf{r}}') \in [1,n]^{2(k+1)}} (X_\pi)_{\tilde{\mathsf{r}}}^{\tilde{\mathsf{r}}'} \mathsf{E}_{\tilde{\mathsf{r}}}^{\tilde{\mathsf{r}}'},$$

where the sum is over tuples of the form $\tilde{\mathsf{r}} = (r_1, \ldots, r_k, n)$, $\tilde{\mathsf{r}}' = (r_{k+1}, \ldots, r_{2k}, n)$ in $[1,n]^{k+1}$, and

$$(\mathsf{X}_\pi)_{\tilde{\mathsf{r}}}^{\tilde{\mathsf{r}}'} = \begin{cases} 1 & \text{if } r_a = r_b \textit{ if and only if } a \text{ and } b \text{ are in the same block of } \pi, \\ 0 & \text{otherwise.} \end{cases} \tag{5.15}$$

Then the argument proving [3, Prop. 3.6] can be easily adapted to show that $\Phi_{k+\frac{1}{2},n}$ is a representation of $\mathsf{P}_{k+\frac{1}{2}}(n)$.

The next theorem describes a basis for the image and the kernel of $\Phi_{k,n}$. Part (a) follows from our work in Sect. 5.1 and is originally due to Jones [22]. The extension to $\Phi_{k+\frac{1}{2},n}$ can be found in [19, Thm. 3.6].

Theorem 5.16 *Assume $n \in \mathbb{Z}_{\geq 1}$ and $\{x_\pi \mid \pi \in \Pi_{2k}\}$ is the orbit basis for* $\mathsf{P}_k(n)$.

(a) *For $k \in \mathbb{Z}_{\geq 1}$, the representation $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}(\mathsf{M}_n^{\otimes k})$ has*

$$\operatorname{im}\Phi_{k,n} = \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k}) = \mathsf{span}_{\mathbb{C}}\{\Phi_{k,n}(x_\pi) \mid \pi \in \Pi_{2k} \textit{ has } \le n \textit{ blocks}\}$$
$$\ker\Phi_{k,n} = \mathsf{span}_{\mathbb{C}}\{x_\pi \mid \pi \in \Pi_{2k} \textit{ has more than } n \textit{ blocks}\}.$$

Consequently, $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ *is isomorphic to* $\mathsf{P}_k(n)$ *for* $n \ge 2k$.

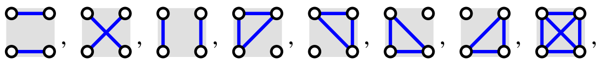
(b) *For* $k \in \mathbb{Z}_{\ge 0}$, *the representation* $\Phi_{k+\frac{1}{2}} : \mathsf{P}_{k+\frac{1}{2}}(n) \to \mathsf{End}(\mathsf{M}_n^{\otimes k})$ *has*

$$\operatorname{im}\Phi_{k+\frac{1}{2},n} = \mathsf{End}_{\mathsf{S}_{n-1}}(\mathsf{M}_n^{\otimes k}) = \mathsf{span}_{\mathbb{C}}\{\Phi_{k+\frac{1}{2},n}(x_\pi) \mid \pi \in \Pi_{2k+1} \textit{ has } \le n \textit{ blocks}\}$$
$$\ker\Phi_{k+\frac{1}{2},n} = \mathsf{span}_{\mathbb{C}}\{x_\pi \mid \pi \in \Pi_{2k+1} \textit{ has more than } n \textit{ blocks}\}.$$

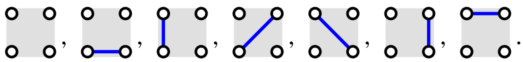
Consequently, $\mathsf{End}_{\mathsf{S}_{n-1}}(\mathsf{M}_n^{\otimes k})$ *is isomorphic to* $\mathsf{P}_{k+\frac{1}{2}}(n)$ *for* $n \ge 2k+1$.

Remark 5.17 The assertion that the map $\Phi_{k,n}$ (resp. $\Phi_{k+\frac{1}{2},n}$) is an isomorphism when $n \ge 2k$ (resp. when $n \ge 2k+1$) holds because set partitions $\pi \in \Pi_{2k}$ (resp. $\pi \in \Pi_{2k+1}$) have no more than n blocks under those assumptions.

Example 5.18 When $k = 2$ and $n = 2$, the image of $\Phi_{2,2} : \mathsf{P}_2(2) \to \mathsf{End}_{\mathsf{S}_2}(\mathsf{M}_2^{\otimes 2})$ is spanned by the images of the following 8 diagrams,



and the kernel is spanned by the following 7 diagrams,



Remark 5.19 Recall from Corollary 3.11 that $\mathsf{B}(\ell, n) = (n\,!)^{-1}\sum_{\sigma\in\mathsf{S}_n}\mathsf{F}(\sigma)^\ell$ for all $\ell \in \mathbb{Z}_{\ge 0}$. When $n = 2$, only the identity element of S_2 has fixed points, and we see that $\mathsf{B}(2k, 2) = \frac{1}{2}(2^{2k}) = 2^{2k-1}$ for all $k \in \frac{1}{2}\mathbb{Z}_{\ge 0}$, in agreement with the values in the first column of Fig. 5. When $n = 3$, there are three transpositions in S_3, each having one fixed point, and two cycles of length 3 that have no fixed points. Thus,

$$\mathsf{B}(2k, 3) = \frac{1}{6}\left(3^{2k} + 3\cdot 1^{2k} + 2\cdot 0^{2k}\right) = \frac{3^{2k-1}+1}{2} \quad \text{for } k \in \frac{1}{2}\mathbb{Z}_{\ge 0}.$$

These values correspond to the numbers in the second column of the table.

5.3 The Kernel of the Surjection $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ When $2k > n$

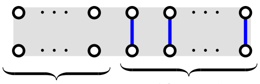
This section is devoted to a description of the kernel of the map $\Phi_{k,n}$ (and also of $\Phi_{k-\frac{1}{2},n}$) when $2k > n$. Toward this purpose, the following orbit basis elements $\mathsf{e}_{k,n}$

k	$\mathsf{B}(2k,2)$	$\mathsf{B}(2k,3)$	$\mathsf{B}(2k,4)$	$\mathsf{B}(2k,5)$	$\mathsf{B}(2k,6)$	$\mathsf{B}(2k,7)$	$\mathsf{B}(2k,8)$	$\mathsf{B}(2k)$
$\frac{1}{2}$	1	1	1	1	1	1	1	1
1	2	2	2	2	2	2	2	2
$1\frac{1}{2}$	4	5	5	5	5	5	5	5
2	8	14	15	15	15	15	15	15
$2\frac{1}{2}$	16	41	51	52	52	52	52	52
3	32	122	187	202	203	203	203	203
$3\frac{1}{2}$	64	365	715	855	876	877	877	877
4	128	1094	2795	3845	4111	4139	4140	4140
$4\frac{1}{2}$	256	3281	11051	18002	20648	21110	21146	21147
5	512	9842	43947	86472	109299	115179	115929	115975
$5\frac{1}{2}$	1024	29525	175275	422005	601492	665479	677359	678570
6	2048	88574	700075	2079475	3403127	4030523	4189550	4213597

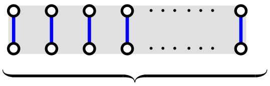
Fig. 5 Table of values of the restricted Bell number $\mathsf{B}(2k, n)$, which equals the dimension of the image of the surjection $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$. The rightmost column gives $\dim(\mathsf{P}_k(n)) = \mathsf{B}(2k)$, the $2k$th (unrestricted) Bell number. Note that column $\mathsf{B}(2k, 5)$ equals the right-hand column of dimensions in the Bratteli diagram $\mathcal{B}(\mathsf{S}_5, \mathsf{S}_4)$ displayed in Fig. 1

for $k, n \in \mathbb{Z}_{\geq 1}$ and $2k > n$ were introduced in [3, Sect. 5.3]:

$\mathsf{e}_{k,n} = \begin{cases} \end{cases}$



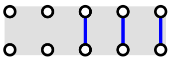
$n+1-k$ $\quad$ $2k-n-1$ $\qquad$ if $n \geq k > n/2$,



k $\qquad$ if $k > n$. (5.20)

$$\mathsf{e}_{k-\frac{1}{2},n} = \mathsf{e}_{k,n}, \qquad \text{if } 2k-1 > n. \tag{5.21}$$

Observe that if $n \geq k > n/2$, then the number of blocks in $\mathsf{e}_{k,n}$ is $|\mathsf{e}_{k,n}| = 2(n+1-k) + 2k - n - 1 = n + 1$, so $\mathsf{e}_{k,n}$ is in the kernel of $\Phi_{k,n}$. For example,

$\mathsf{e}_{4\frac{1}{2},6} = \mathsf{e}_{5,6} =$ 

has $|\mathsf{e}_{5,6}| = 7$ blocks. The elements $\mathsf{e}_{k,n}$ for $k \leq 5$ and $n \leq 9$ are displayed in Fig. 6.

Theorem 5.22 ([3, Thms. 5.6 and 5.9]) *Assume $k, n \in \mathbb{Z}_{\geq 1}$ and $2k > n$.*

(a) *The orbit basis element $\mathsf{e}_{k,n}$ in (5.20) is an essential idempotent such that $(\mathsf{e}_{k,n})^2 = c_{k,n}\mathsf{e}_{k,n}$, where*

$$c_{k,n} = \begin{cases} (-1)^{n+1-k}\,(n+1-k)! & \text{if } n \geq k > n/2, \\ 1 & \text{if } k > n. \end{cases} \tag{5.23}$$

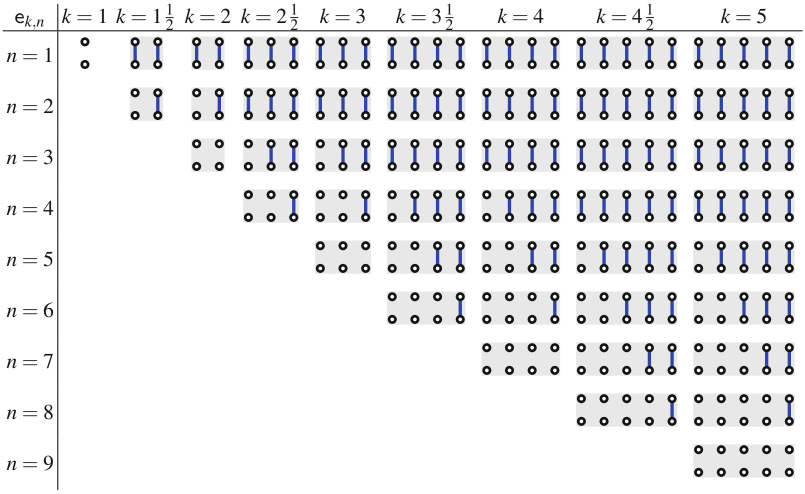


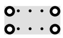
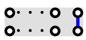
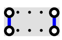
Fig. 6 The essential idempotent $\mathsf{e}_{k,n}$ for $k \leq 5$ and $n \leq 9$. When $n < 2k$, the kernel of $\Phi_{k,n}$ equals the principal ideal $\langle \mathsf{e}_{k,n} \rangle$

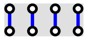
(b) *The kernel of the representation* $\Phi_{k,n}$ *is the ideal of* $\mathsf{P}_k(n)$ *generated by* $\mathsf{e}_{k,n}$ *when* $2k > n$.
(c) *The kernel of the representation* $\Phi_{k-\frac{1}{2},n}$ *is the ideal of* $\mathsf{P}_{k-\frac{1}{2}}(n)$ *generated by* $\mathsf{e}_{k-\frac{1}{2},n} = \mathsf{e}_{k,n}$ *when* $2k > n + 1$.

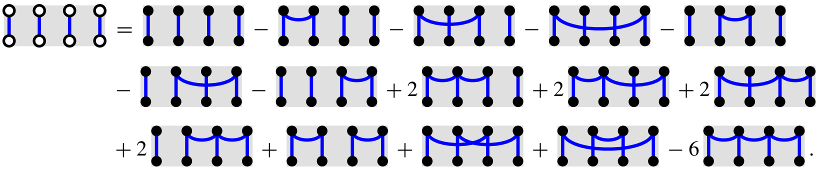
Remark 5.24 For a fixed value of n, the first time the kernel is nonzero is when $k = \frac{1}{2}(n+1)$ (i.e., when $n = 2k - 1$). This is the first entry in each row in the table in Fig. 6. For that particular value of n, $(\mathsf{e}_{k,2k-1})^2 = (-1)^k/k!\, \mathsf{e}_{k,2k-1}$ by Theorem 5.22(a).

The expression for $\mathsf{e}_{k,n}$ in the diagram basis is given by

$$\mathsf{e}_{k,n} = \sum_{\rho \in \Pi_{2k},\ \pi_{k,n} \preceq \rho} \mu_{2k}(\pi_{k,n}, \rho) d_\rho, \tag{5.25}$$

where $\pi_{k,n}$ is the set partition of $[1, 2k]$ corresponding to $\mathsf{e}_{k,n}$. When $k = \frac{1}{2}(n+1)$ and n is odd, $\mathsf{e}_{k,n} =$ , and all ρ in Π_{2k} occur in the expression for $\mathsf{e}_{k,n}$. When $k = \frac{1}{2}(n+1)$ and n is even, $\mathsf{e}_{k,n} =$ , and all ρ in Π_{2k-1} occur in the expression for $\mathsf{e}_{k,n}$. If $k > n$, then $\mathsf{e}_{k,n} =$ , and the set partitions $\rho \in \Pi_{2k}$ that occur in the expression for $\mathsf{e}_{k,n}$ correspond to the coarsenings of the k columns of the diagram $\pi_{k,n}$. There are Bell number $\mathsf{B}(k)$ such terms. In each case, the integer coefficients $\mu_{2k}(\pi_{k,n}, \rho)$ can be computed using (4.9).

Remark 5.26 The element $\mathsf{e}_{4,3} =$  generates the kernel of the surjection $\Phi_{4,3} : \mathsf{P}_4(3) \to \mathsf{End}_{\mathsf{S}_3}(\mathsf{M}_3^{\otimes 4})$. Below is the diagram basis expansion for $\mathsf{e}_{4,3}$ using the Möbius formulas in (4.8) and (4.9). The diagram basis expansion for $\mathsf{e}_{2,3}$ can be found in (4.10).

$\mathsf{e}_{4,3} =$ 

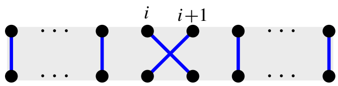
There are $15 = \mathsf{B}(4)$ diagram basis elements in the expression for orbit element $\mathsf{e}_{k,n}$.

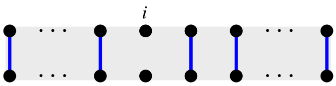
6 The Fundamental Theorems of Invariant Theory for S_n

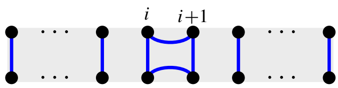
Section 5.2 gives the explicit construction of the algebra homomorphism $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ and shows that the partition algebra generates the tensor invariants of the symmetric group S_n. The First Fundamental Theorem of Invariant Theory for S_n says that the partition algebra generates *all* tensor invariants of S_n on $\mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k}) \cong \left(\mathsf{M}_n^{\otimes 2k}\right)^{\mathsf{S}_n}$, as $\Phi_{k,n}$ is a surjection for all k, n. A more precise statement is the following:

Theorem 6.1 ([22]) (First Fundamental Theorem of Invariant Theory for S_n) *For all $k, n \in \mathbb{Z}_{\geq 1}$, $\Phi_{k,n} : \mathsf{P}_k(n) \to \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ is a surjective algebra homomorphism, and when $n \geq 2k$, $\Phi_{k,n}$ is an isomorphism, so $\mathsf{P}_k(n) \cong \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ when $n \geq 2k$.*

For $k \in \mathbb{Z}_{\geq 1}$, the partition algebra $\mathsf{P}_k(n)$ has a presentation by the generators

$\mathfrak{s}_i =$  $1 \leq i \leq k-1,$ (6.2)

$\mathfrak{p}_i = \dfrac{1}{n}$  $1 \leq i \leq k,$ (6.3)

$\mathfrak{b}_i =$  $1 \leq i \leq k-1,$ (6.4)

and the relations in the next result. Additional results on presentations for partition algebras can be found in [11] and the references therein, and in [12] which adopts a Jucys–Murphy element point of view.

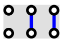
Theorem 6.5 [19, Thm. 1.11] *Assume $k \in \mathbb{Z}_{\geq 1}$, and set $\mathfrak{p}_{i+\frac{1}{2}} = \mathfrak{b}_i$ $(1 \leq i \leq k-1)$. Then $\mathsf{P}_k(n)$ has a presentation as a unital associative algebra by generators $\mathfrak{s}_i$ $(1 \leq i \leq k-1)$, $\mathfrak{p}_\ell$ $(\ell \in \frac{1}{2}\mathbb{Z}_{\geq 1},\ 1 \leq \ell \leq k)$, and the following relations:*

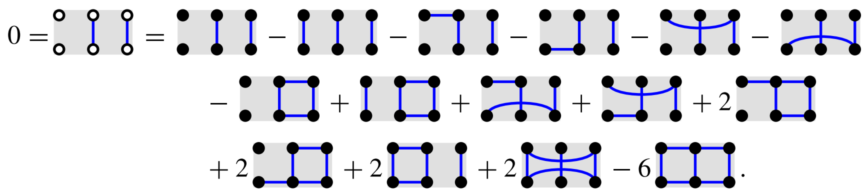
(a) $\mathfrak{s}_i^2 = \mathsf{I}_k, \quad \mathfrak{s}_i\mathfrak{s}_j = \mathfrak{s}_j\mathfrak{s}_i \quad (|i-j| > 1), \quad \mathfrak{s}_i\mathfrak{s}_{i+1}\mathfrak{s}_i = \mathfrak{s}_{i+1}\mathfrak{s}_i\mathfrak{s}_{i+1} \quad (1 \leq i \leq k-2);$
(b) $\mathfrak{p}_\ell^2 = \mathfrak{p}_\ell, \quad \mathfrak{p}_\ell\mathfrak{p}_m = \mathfrak{p}_m\mathfrak{p}_\ell \quad (m \neq \ell \pm \frac{1}{2}), \quad \mathfrak{p}_\ell\mathfrak{p}_{\ell\pm\frac{1}{2}}\mathfrak{p}_\ell = \mathfrak{p}_\ell \quad (\mathfrak{p}_{\frac{1}{2}} := \mathsf{I}_k =: \mathfrak{p}_{k+\frac{1}{2}});$
(c) $\mathfrak{s}_i\mathfrak{p}_i\mathfrak{p}_{i+1} = \mathfrak{p}_i\mathfrak{p}_{i+1}, \quad \mathfrak{s}_i\mathfrak{p}_i\mathfrak{s}_i = \mathfrak{p}_{i+1},$
$\mathfrak{s}_i\mathfrak{p}_{i+\frac{1}{2}} = \mathfrak{p}_{i+\frac{1}{2}}\,\mathfrak{s}_i = \mathfrak{p}_{i+\frac{1}{2}} \quad (1 \leq i \leq k-1),$
$\mathfrak{s}_i\mathfrak{s}_{i+1}\mathfrak{p}_{i+\frac{1}{2}}\mathfrak{s}_{i+1}\mathfrak{s}_i = \mathfrak{p}_{i+\frac{3}{2}} \quad (1 \leq i \leq k-2),$
$\mathfrak{s}_i\mathfrak{p}_\ell = \mathfrak{p}_\ell\mathfrak{s}_i \quad (\ell \neq i-\frac{1}{2}, i, i+\frac{1}{2}, i+1, i+\frac{3}{2}).$

Remark 6.6 It is easily seen from the relations that $\mathsf{P}_k(n)$ is generated by the elements $\mathfrak{p}_1$, $\mathfrak{b}_1 = \mathfrak{p}_{1+\frac{1}{2}}$, and $\mathfrak{s}_i$ $(1 \leq i \leq k-1)$.

Theorems 5.6 and 5.8 of [3] prove that $\mathsf{e}_{k,n}$ is an essential idempotent that generates the kernel of $\Phi_{k,n}$ as a two-sided ideal. Moreover, Theorem 5.15 of [3] shows that the kernel of $\Phi_{k,n}$ is generated as a two-sided ideal by the embedded image $\mathsf{e}_{n,n} \otimes$ ()$^{\otimes(k-n)}$ (the diagram of $\mathsf{e}_{n,n}$ with $k-n$ vertical edges  juxtaposed to its right) of the essential idempotent $\mathsf{e}_{n,n}$ in $\mathsf{P}_k(n)$ for all $k \geq n$. By [3, Remark 5.20], $\ker \Phi_{k,n}$ cannot be generated by $\mathsf{e}_{\ell,n} \otimes$ ()$^{\otimes(k-\ell)}$ for any ℓ such that $k \geq n > \ell \geq \frac{1}{2}(n+1)$. Identifying $\mathsf{e}_{n,n}$ with its image in $\mathsf{P}_k(n)$ for $k \geq n$, we have

Theorem 6.7 [3, Thm. 5.19] *(Second Fundamental Theorem of Invariant Theory for S_n) For all $k, n \in \mathbb{Z}_{\geq 1}$, $\mathsf{im}\,\Phi_{k,n} = \mathsf{End}_{\mathsf{S}_n}(\mathsf{M}_n^{\otimes k})$ is generated by the partition algebra generators and relations in Theorem 6.5(a)–(c) together with the one additional relation $\mathsf{e}_{k,n} = 0$ in the case that $2k > n$. When $k \geq n$, the relation $\mathsf{e}_{k,n} = 0$ can be replaced with $\mathsf{e}_{n,n} = 0$.*

Example 6.8 The kernel of $\Phi_{3,3} : \mathsf{P}_3(3) \to \mathsf{End}_{\mathsf{S}_3}(\mathsf{M}_3^{\otimes 3})$ is generated by $\mathsf{e}_{3,3} =$ , and the embedded element $\mathsf{e}_{3,3} \otimes$ ()$^{\otimes(k-3)}$ principally generates the kernel for $\mathsf{P}_k(n)$ for all $k \geq 3$. The image, $\mathsf{im}\,\Phi_{3,3} \cong \mathsf{P}_3(3)/\mathsf{ker}\Phi_{3,3}$, is generated by the partition algebra $\mathsf{P}_3(3)$ with the additional dependence relation



This dependence relation is analogous to the one that comes from setting the kernel generator $\sum_{\sigma\in\mathsf{S}_{n+1}}(-1)^{\mathsf{sgn}(\sigma)}\sigma$ of the surjection $\mathbb{F}\mathsf{S}_k \to \mathsf{End}_{\mathsf{GL}_n}(\mathsf{V}^{\otimes k})$ $(\mathsf{V} = \mathbb{F}^n)$ equal to 0 in the Second Fundamental Theorem of Invariant Theory for GL_n.

Acknowledgements The authors thank the referee for a careful proofreading and useful suggestions.

References

1. J. Barnes, G. Benkart, T. Halverson, McKay centralizer algebras. Proc. Lond. Math. Soc. **112**(2), 375–414 (2016)
2. G. Benkart, M. Chakrabarti, T. Halverson, R. Leduc, C. Lee, J. Stroomer, Tensor product representations of general linear groups and their connections with Brauer algebras. J. Algebra **166**(3), 529–567 (1994)
3. G. Benkart, T. Halverson, Partition algebras $\mathsf{P}_k(n)$ with $2k>n$ and the fundamental theorems of invariant theory for the symmetric group S_n, J. Lond. Math. Soc. in press. arXiv:1707.1410
4. G. Benkart, T. Halverson, N. Harman, Dimensions of irreducible modules for partition algebras and tensor power multiplicities for symmetric and alternating groups. J. Algebraic Combin. **46**(1), 77–108 (2017), arXiv:1605.06543
5. G. Benkart, D. Moon, Walks on graphs and their connections with tensor invariants and centralizer algebras, J. Algebra **509**, 1–39 (2018). arXiv:1610.07837
6. C. Bowman, M. DeVisscher, R. Orellana, The partition algebra and the Kronecker coefficients. Discrete Math. Theor. Comput. Sci. Proc. AS 321–332 (2013)
7. C. Bowman, M. DeVisscher, R. Orellana, The partition algebra and the Kronecker coefficients. Trans. Amer. Math. Soc. **367**, 3647–3667 (2015)
8. C. Bowman, J. Enyang, and F.W. Goodman, The cellular second fundamental theorem of invariant theory for classical groups, arXiv:1610.09009
9. R. Brauer, On algebras which are connected with the semisimple continuous groups. Ann. of Math. **38**(4), 857–872 (1937)
10. W. Chen, E. Deng, R. Du, R. Stanley, C. Yan, Crossings and nestings of matchings and partitions. Trans. Amer. Math. Soc. **359**, 1555–1575 (2007)
11. J. East, Presentations for (singular) partition monoids: a new approach, in *Mathematical Proceedings of the Cambridge Philosophical Society*, First View https://doi.org/10.1017/S030500411700069X, Published online: (11 October 2017), pp. 1–14
12. J. Enyang, Jucys-Murphy elements and a presentation for partition algebras. J. Algebraic Combin. **37**, 401–454 (2013)
13. J. Enyang, A seminormal form for partition algebras. J. Combin. Theory Ser. A **120**, 1737–1785 (2013)
14. J. Farina, T. Halverson, Character orthogonality for the partition algebra and fixed points of permutations. Adv. Appl. Math. **31**, 113–131 (2003)
15. W. Fulton, J. Harris, *Representation Theory, A First Course*. Graduate Texts in Mathematics, vol. 129 (Springer, New York, 1991)
16. R. Goodman, N.R. Wallach, *Representations and Invariants of the Classical Groups*. Encyclopedia of Mathematics and Its Applications, vol. 68 (Cambridge University Press, Cambridge, 1998); 3rd corrected printing, Cambridge University Press (2003)
17. T. Halverson, Characters of the partition algebras. J. Algebra **238**, 502–533 (2001)
18. T. Halverson, T. Lewandowski, RSK insertion for set partitions and diagram algebras, Electron. J. Combin. vol. 11, no. 2 (2004/2006), 24 pp (Research Paper 24)
19. T. Halverson, A. Ram, Partition algebras. European J. Combin. **26**, 869–921 (2005)
20. J. Hu, Z. Xiao, On tensor spaces for Birman-Murakami-Wenzl algebras. J. Algebra **324**, 2893–2922 (2010)
21. G. James, A. Kerber, *The Representation Theory of the Symmetric Group*. Encyclopedia of Mathematics and Its Applications, vol. 16 (Addison-Wesley Publishing Co., Reading, Mass, 1981)
22. V.F.R. Jones, The Potts model and the symmetric group, in *Subfactors: Proceedings of the Taniguchi Symposium on Operator Algebras (Kyuzeso, 1993)* (World Scientific Publishing, River Edge, NJ, 1994), pp. 259–267
23. G. Lehrer, R.B. Zhang, The second fundamental theorem of invariant theory for the orthogonal group. Ann. of Math. **176**, 2031–2054 (2012)
24. G. Lehrer, R.B. Zhang, The Brauer category and invariant theory. J. Eur. Math. Soc. **17**, 2311–2351 (2015)

25. I.G. Macdonald, *Symmetric Functions and Hall Polynomials*, 2nd edn. Oxford Classic Texts in the Physical Sciences (The Clarendon Press, Oxford University Press, New York, 2015)
26. P. Martin, Representations of graph Temperley-Lieb algebras. Publ. Res. Inst. Math. Sci. **26**(3), 485–503 (1990)
27. P. Martin, Temperley-Lieb algebras for non-planar statistical mechanics– the partition algebra construction. J. Knot Theory Ramifications **3**, 51–82 (1994)
28. P. Martin, The structure of the partition algebra. J. Algebra **183**, 319–358 (1996)
29. P. Martin, G. Rollet, The Potts model representation and a Robinson-Schensted correspondence for the partition algebra. Compositio Math. **112**, 237–254 (1998)
30. A. Morales, I. Pak, G. Panova, Hook formulas for skew shapes I. q-analogues and bijections. J. Combin. Theory Ser. A **154**, 350–405 (2018)
31. R. Orellana, M. Zabrocki, *Symmetric group characters as symmetric functions*, arXiv:1605.06672v3
32. R. Orellana, M. Zabrocki, *Products of characters of the symmetric group*, arXiv:1709.08098
33. M. Rubey, B. Westbury, A combinatorial approach to classical representation theory, arXiv:1408.3592
34. M. Rubey, B. Westbury, Combinatorics of symplectic invariant tensors, in *Proceedings of FPSAC 2015, Discrete Mathematics & Theoretical Computer Science Proceedings of the Association; Discrete Mathematics & Theoretical Computer Science, Nancy* (2015), pp. 285–296
35. B.E. Sagan, *The symmetric group. Representations, Combinatorial Algorithms, and Symmetric Functions*, 2 edn. Graduate Texts in Mathematics, vol. 203 (Springer, New York, 2001), pp. xvi+238
36. R.P. Stanley, *Enumerative Combinatorics*, vol. 1 (Cambridge, England, 1997)
37. R.P. Stanley, *Enumerative Combinatorics*, vol. 2 (Cambridge, England, 1999)

Affine Grassmannians and Hessenberg Schubert Cells

Linda Chen and Julianna Tymoczko

Abstract We give an overview of the linear algebra, geometry, and combinatorics of affine Grassmannians along the lines of Fulton's *Young Tableaux* for classical Grassmannians. We discuss geometric and linear algebraic aspects of the decomposition of the affine Grassmannian into affine Schubert cells in terms of coset representatives and linear models. We describe (Grassmannian) Hessenberg Schubert cells and show that every affine Schubert cell can be realized as a Hessenberg Schubert cell in a complete flag variety and as a Grassmannian Hessenberg Schubert cell in a finite Grassmannian.

1 Introduction

Let $\mathrm{Gr}_n = GL_n(\mathbb{C}((t)))/GL_n(\mathbb{C}[[t]])$ denote the affine Grassmannian of type A_{n-1}, where $\mathbb{C}((t))$ is the ring of formal Laurent series and $\mathbb{C}[[t]]$ is the ring of formal power series. It is an infinite dimensional algebraic variety that is a central object in algebraic combinatorics, algebraic geometry, and geometric representation theory. Some introductions to the affine Grassmannian in the literature include that by Lam, Lapointe, Morse, and Shimozono [16], which gives an overview of the Schubert calculus and tableaux combinatorics of affine Grassmannians, and that by Zhu [27], which describes connections to Kac–Moody groups, the moduli of vector bundles on curves, the Langlands program, and the geometric Satake equivalence.

LC was partially supported by NSF Grant DMS-1101625 and Simons Foundation Collaboration Grant 524354. JT was partially supported by NSF Grant DMS-1362855.

L. Chen (✉)
Department of Mathematics and Statistics, Swarthmore College,
Swarthmore, PA 19081, USA
e-mail: lchen@swarthmore.edu

J. Tymoczko
Department of Mathematics and Statistics, Smith College,
Northampton, MA 01063, USA
e-mail: jtymoczko@smith.edu

H. Barcelo et al. (eds.), *Recent Trends in Algebraic Combinatorics*,
Association for Women in Mathematics Series 16,
https://doi.org/10.1007/978-3-030-05141-9_2

The geometry of the affine Grassmannian has also been extensively studied. The homology and cohomology rings of Gr_n were computed by Bott as a subring and quotient ring of the ring of symmetric functions. A basis consisting of Schubert varieties and opposite Schubert varieties was studied by Kostant and Kumar using the nilHecke ring [14], and Lam identified them as k-Schur and dual k-Schur functions, which arise also in the study of Macdonald polynomials [15]. Lapointe and Morse gave explicit combinatorial maps between k-Schur functions in the homology of the affine Grassmannian and the quantum cohomology rings of the classical Grassmannian [19].

In this paper, we give a basic and explicit linear algebraic description of the geometry of *affine Schubert cells*, which are indexed by *affine Grassmannian permutations*. We show how this gives rise to some of the known combinatorics, as outlined in [16, 17]. In particular, we describe several equivalent formulations of affine Grassmannian permutations and their corresponding cells, for instance in terms of windows, skylines, the geometric linear model, the coset interpretation, cores, and partitions, extending analogous formulations in the case of the classical Grassmannian.

We also study Hessenberg varieties, which form a family of subvarieties of the complete flag variety defined by two parameters. Here, the complete flag variety $Fl(n)$ is the space of flags $V_\bullet = \{V_1 \subseteq V_2 \subseteq \cdots \subseteq V_n = \mathbb{C}^n\}$ where each V_i is an i-dimensional subspace. Given a nilpotent linear operator X and a nondecreasing step function $h : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$, the Hessenberg variety $\mathcal{H}(X, h)$ consists of the flags $V_\bullet$ such that $XV_i \subseteq V_{h(i)}$ for each i.

Hessenberg varieties arise in many contexts. They were originally defined by De Mari, Procesi, and Shayman as a generalization of spaces that appear in numerical analysis [5]. When $h(i) = i$ for all i, they are known as *Springer fibers*; the top-dimensional cohomology of Springer fibers carries a representation of the symmetric group, and all irreducible representations can be obtained bijectively by varying over the nilpotent conjugacy classes of X [11, 23]. When X is regular nilpotent (i.e., has a single Jordan block) and $h(i) = i + 1$ for all $i \leq n - 1$, the Hessenberg varieties $\mathcal{H}(X, h)$ can be used to obtain the quantum cohomology of the flag variety [13, 22]. More generally, when X is regular nilpotent, Hessenberg varieties are geometrically associated with affine Grassmannians. Peterson proved an explicit isomorphism between a localized quantum cohomology ring of Grassmannians and a localized equivariant cohomology ring of the affine Grassmannian [20]. The geometry of Hessenberg varieties is quite mysterious: we only have limited information about fundamental properties such as singularities of their components, even for Springer fibers [7, 8, 10, 12].

We describe *Hessenberg Schubert cells* which are the intersection of a Hessenberg variety with a Schubert cell in $Fl(n)$; their projections to a Grassmannian $Gr(k, n)$ are called *Grassmannian Hessenberg Schubert cells*. We show that every affine Schubert cell can be realized as a Grassmannian Hessenberg Schubert cell. We do this in two ways and show that for one of them, the affine Schubert cell is in fact isomorphic to the original Hessenberg Schubert cell. This result does three things: (1) it further motivates and contextualizes ongoing work on Hessenberg Schubert calculus

[2, 3, 6], (2) it turns affine Schubert calculus into a finite rather than infinite problem, and (3) it introduces the study of Grassmannian Hessenberg varieties.

For our purposes in this survey, we work with $GL_n(\mathbb{C}[[t]])$ though $SL_n(\mathbb{C}[[t]])$ is also common in the literature. The substantive differences between the two cases are slight.

2 The Classical Grassmannian

The Grassmannian is a central object in algebraic geometry, combinatorics, and representation theory. A general introduction to the geometry of Grassmannians and flag varieties and associated combinatorics of Young tableaux can be found in [9]. In this section, we review the features of the Grassmannian that are most useful to understand when studying the affine Grassmannian. The Grassmannian $Gr(k, n)$ of k-planes in $\mathbb{C}^n$ can be described as the quotient $GL_n(\mathbb{C})/P$ where P is a maximal parabolic subgroup of invertible block upper-triangular matrices of the form

$$P = \left(\begin{array}{c|c} * & * \\ \hline 0 & * \end{array}\right)$$

where the entries in the blocks labeled $*$ are arbitrary (subject to the condition that the matrix is invertible), and the zero block is $(n-k) \times k$. In other words,

$$P = \{(m_{ij}) \in GL_n(\mathbb{C}) : m_{ij} = 0 \text{ if } i > k \text{ and } j < n-k\}. \tag{1}$$

The Weyl group in type A_{n-1} is the symmetric group $W = S_n$ generated by the simple reflections $s_1, s_2, \ldots, s_{n-1}$ subject to the relations

$$\begin{gathered} s_i^2 = 1 \text{ for all } i \\ s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1} \text{ if } i = 1, 2, \ldots, n-2, \text{ and} \\ s_i s_j = s_j s_i \text{ if } |i-j| > 1. \end{gathered}$$

Given integers k and n such that $1 \leq k \leq n$, the *Grassmannian permutations with (possible) descent at* k are the permutations $\sigma \in S_n$ such that $\sigma(1) < \cdots < \sigma(k)$ and $\sigma(k+1) < \cdots < \sigma(n)$. In fact, Grassmannian permutations with descent at k are the minimal coset representatives for W/W_P, where W_P is the subgroup of W generated by all simple reflections except s_k, namely $W_P = \langle s_j : j \neq k \rangle \cong S_k \times S_{n-k}$.

We now give concrete descriptions of these objects.

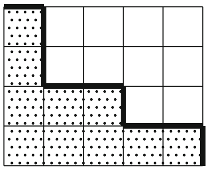
2.1 Combinatorial Description of Grassmannian Permutations

Grassmannian permutations have combinatorial descriptions in terms of partitions and bit sequences, as follows. Given a Grassmannian permutation σ with (possible)

descent at k, we associate a partition $\lambda(\sigma) = (\sigma(k) - k, \ldots, \sigma(1) - 1)$, which by construction lies inside a $k \times (n-k)$ rectangle. Moreover, since σ is a minimal coset representative for its coset in $W/W_P \cong S_n/(S_k \times S_{n-k})$, these cosets can be described combinatorially via a set on which W acts transitively with stabilizer (at a point) $\langle s_j : j \neq k \rangle$.

Consider the set of n-bit sequences with exactly k zeros. For each i let the simple reflection s_i act on an n-bit sequence by exchanging the elements in positions i and $i+1$. This set corresponds bijectively with the Grassmannian permutations. The permutation σ is recovered from the bit sequence by letting the ones be indexed by $\sigma(1), \ldots, \sigma(k)$ and the zeros by $\sigma(k+1), \ldots, \sigma(n)$. We can construct the partition from the bit sequence as follows: starting from the northwest corner of a $k \times (n-k)$ rectangle, traverse the sequence by moving south for each 1 and east for every 0.

Example 2.1. *For $Gr(4,9)$, the permutation $\sigma = 236914578$ is a Grassmannian permutation with descent at $k = 4$. The associated partition $\lambda(\sigma)$ is $(5,3,1,1)$ and the bit sequence is* 011001001.



2.2 Linear Model

Linear algebra is at the heart of traditional Schubert calculus. The Grassmannian $Gr(k,n)$ is often described as the collection of k-dimensional planes in $\mathbb{C}^n$. Given a k-dimensional subspace, choose k spanning vectors and write them as the columns of an $n \times k$ matrix. Then use Gaussian elimination on the column vectors to obtain a normal form for the k-plane. For instance, each element of $Gr(2,4)$ appears exactly once in the following list of matrices, where $*$ denotes free entries in $\mathbb{C}$:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ * & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ * & 0 \\ * & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ * & * \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ * & 0 \\ 0 & 1 \\ * & * \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ * & * \\ * & * \end{pmatrix}.$$

2.3 Decomposition into Schubert Cells

Let B denote the Borel subgroup of $GL_n(\mathbb{C})$ consisting of invertible $n \times n$ upper-triangular matrices. Then the Grassmannian $Gr(k,n)$ has a decomposition

$$Gr(k,n) = \bigcup B\sigma P/P$$

into a disjoint union of *Schubert cells* $B\sigma P/P$ indexed by Grassmannian permutations σ with possible descent at k. Similarly, when B^- is the opposite Borel subgroup consisting of invertible $n \times n$ lower-triangular matrices, there is a decomposition

$$Gr(k,n) = \bigcup B^-\sigma P/P$$

into a disjoint union of *opposite Schubert cells* $B^-\sigma P/P$.

In order to be compatible with the standard conventions for the affine Grassmannian (as in the next sections), we consider the opposite Schubert cell corresponding to a partition λ in a $k \times (n-k)$ rectangle (or equivalently, a Grassmannian permutation with descent at k) [9, §9.4]. When written as in Sect. 2.2, the free entries of the matrices in the opposite Schubert cell corresponding to σ form the **transpose dual** of λ, where the *dual partition* $(n-k-\lambda_k, \ldots, n-k-\lambda_1)$ is the complement of λ and the transpose (or conjugate) means we take columns rather than rows of the partition (corresponding to our use of column vectors rather than the row vectors Fulton uses [9]).

To see how a k-plane corresponds to a coset aP, we complete each $n \times k$ matrix in column-reduced echelon form to an $n \times n$ matrix so that the $n-k$ rows without pivots in the first k columns restrict to an $(n-k) \times (n-k)$ identity matrix in the last $n-k$ columns, and so that all other entries in the last $n-k$ columns are zero. With this description, an opposite Schubert cell in $Gr(k,n)$ corresponds to a collection of $n \times k$ matrices that are in column-reduced echelon form and that have the identity matrix in a fixed $k \times k$ minor.

Example 2.2. *For the permutation* $\sigma = 236914578$ *in* $Gr(4,9)$*, the opposite Schubert cell is given below for the* 4*-plane. To the right we show the minimal coset representative for* aP*. Note that the free entries form the partition* $(3,3,2,2)$*, which is the transpose dual of the partition* $\lambda(\sigma) = (5,3,1,1)$.

$$\begin{pmatrix} 0&0&0&0\\ 1&0&0&0\\ 0&1&0&0\\ *&*&0&0\\ *&*&0&0\\ 0&0&1&0\\ *&*&*&0\\ *&*&*&0\\ 0&0&0&1 \end{pmatrix} \qquad \begin{pmatrix} 0&0&0&0&1&0&0&0&0\\ 1&0&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0\\ *&*&0&0&0&1&0&0&0\\ *&*&0&0&0&0&1&0&0\\ 0&0&1&0&0&0&0&0&0\\ *&*&*&0&0&0&0&1&0\\ *&*&*&0&0&0&0&0&1\\ 0&0&0&1&0&0&0&0&0 \end{pmatrix}$$

3 Affine Grassmannian Permutations

In this section, we give several descriptions of affine Grassmannian permutations which, analogously to Grassmannian permutations, index affine Schubert cells. We

use these permutations to describe various aspects of the affine Grassmannian, including coset descriptions and linear models.

The group of affine permutations $\widetilde{S}_n$ of type $\widetilde{A}_{n-1}$ is the group $\langle s_0, s_1, s_2, \dots, s_{n-1}\rangle$ generated by the simple reflections $s_0, s_1, s_2, \dots, s_{n-1}$ subject to the relations

$$s_i^2 = 1 \text{ for all } i$$
$$s_i s_{i+1(\text{mod } n)} s_i = s_{i+1(\text{mod } n)} s_i s_{i+1(\text{mod } n)} \text{ if } i = 0, 1, 2, \dots, n-1, \text{ and}$$
$$s_i s_j = s_j s_i \text{ if } |i - j| > 1.$$

For $Gr(k, n)$, the Grassmannian permutations with descent at k are exactly the elements of minimal length in their cosets in $S_n/(S_k \times S_{n-k})$. Compare to the set of affine Grassmannian permutations, which are the elements of minimal length in their cosets in $\widetilde{S}_n/S_n$.

We now give more explicit combinatorial descriptions of the affine Grassmannian permutations, beginning with a parallel to bit strings in the Grassmannian case. The affine permutation group $\widetilde{S}_n$ can be described as the set of bijections $w : \mathbb{Z} \to \mathbb{Z}$ such that

- $w(i + n) = w(i) + n$ for all $i \in \mathbb{Z}$, and
- $\displaystyle\sum_{i=1}^{n}(w(i) - i) = 0.$

For $0 \le i \le n$, the simple reflections s_i can be defined by

- $s_i(na + i) = na + i + 1$ and $s_i(na + i + 1) = na + i$, and
- s_i acts as the identity on all other integers.

Because of the characterization that $w(i + n) = w(i) + n$, we sometimes denote an affine permutation $w \in \widetilde{S}_n$ by its uniquely determined *window*

$$[w(1), \dots, w(n)].$$

In window notation, the reflection s_i exchanges the entries in positions i and $i + 1$ for $1 \le i \le n - 1$ and $ws_0 = [w(n) - n, w(2), \dots, w(1) + n]$.

Moreover, for $1 \le i \le n$, we can divide to express $w(i) = nk_i + \sigma(i)$ with $1 \le \sigma(i) \le n$. Inspecting the description of the window of w, we see that σ is a permutation of $\{1, \dots, n\}$ and $\sum_{i=1}^{n} k_i = 0$. It is sometimes convenient to write the window of w as

$$[w(1), \dots, w(n)] = n[k_1, \dots, k_n] + [\sigma(1), \dots, \sigma(n)].$$

We say a permutation w is an *affine Grassmannian permutation* if its window is increasing, namely $w(1) < \cdots < w(n)$. This is the minimal length element in the coset $wS_n \in \widetilde{S}_n/S_n$. Given an affine Grassmannian permutation, we reorder the k_i according to σ to define the *offset sequence* $d(w) = (d_1, \dots, d_n)$. In other words, for each i we have $d_{\sigma(i)} = k_i$ or equivalently $d_i = k_{\sigma^{-1}(i)}$.

Thus given an affine Grassmannian permutation w, we obtain an element $d(w)$ of the set

$$\mathcal{S} = \left\{ (d_1, d_2, \ldots, d_n) \in \mathbb{Z}^n : \sum_{i=1}^{n} d_i = 0 \right\}.$$

Conversely, given such an n-tuple $d = (d_1, d_2, \ldots, d_n)$, we obtain a unique affine Grassmannian permutation whose window is the set $\{nd_i + i\}$ taken in increasing order.

Example 3.1. *Consider the permutation* $w = s_0 s_2 s_1 s_0 \in \widetilde{S}_3$. *Its window is* $[-1, 0, 7]$*, which can be written* $[-1, 0, 7] = 3[-1, -1, 2] + [2, 3, 1]$ *so that* $k = [-1, -1, 2]$ *and* $d(w) = (2, -1, -1)$. *The set* $\{nd_i + i\} = \{7, -1, 0\}$ *taken in increasing order recovers the window* $[-1, 0, 7]$.

Example 3.2. *Consider the permutation* $w = s_2 s_0 s_1 s_2 s_1 s_0 \in \widetilde{S}_3$. *Its window is* $[-4, 3, 7]$*, which can be written* $[-4, 3, 7] = 3[-2, 0, 2] + [2, 3, 1]$ *so that* $k = [-2, 0, 2]$ *and* $d(w) = [2, -2, 0]$. *Then the set* $\{nd_i + i\} = \{7, -4, 3\}$ *taken in increasing order recovers the window* $[-4, 3, 7]$.

If $i = 1, \ldots, n-1$ then since $s_i(nd_i + i) = nd_i + i + 1$ and $s_i(nd_{i+1} + i + 1) = nd_{i+1} + i$, the simple reflection s_i acts on an element in $\mathcal{S}$ by exchanging the entries d_i and d_{i+1} . Since $s_0(nd_1 + 1) = nd_1 = n(d_1 - 1) + n$ and $s_0(nd_n + n) = nd_n + n + 1 = n(d_n + 1) + 1$, the simple reflection s_0 acts on an element d in $\mathcal{S}$ by sending

$$(d_1, \ldots, d_n) \mapsto (d_n + 1, d_2, \ldots, d_{n-1}, d_1 - 1). \tag{2}$$

With this identification, $\widetilde{S}_n$ acts transitively on $\mathcal{S}$ and the stabilizer of $(0, 0, \ldots, 0)$ is S_n, so $\mathcal{S}$ is in bijection with the set of affine Grassmannian permutations.

3.1 Coset Description for the Affine Grassmannian

We now investigate the elements of the affine Grassmannian as cosets of $GL_n(\mathbb{C}[[t]])$. By analogy with the classical Grassmannian in Sect. 2, we describe the column echelon form as an element of $GL_n(\mathbb{C}[[t]])$. We use these tools later in Sect. 4 to give a stratification of the affine Grassmannian by affine Schubert cells, which are double cosets $IwGL_n(\mathbb{C}[[t]])$ where the *Iwahori subgroup* $I \subseteq GL_n(\mathbb{C}[[t]])$ is an analogue of the Borel subgroup and w is indexed by the set of affine Grassmannian permutations.

For every $g \in GL_n(\mathbb{C}((t)))$, we construct a column echelon matrix by applying a version of Gaussian elimination on elements of $GL_n(\mathbb{C}[[t]])$. Since there exists a unique column echelon matrix in each coset $gGL_n(\mathbb{C}[[t]])$, the set of column echelon matrices give a set of representatives for the cosets $gGL_n(\mathbb{C}[[t]])$.

Note that in linear algebra over $\mathbb{C}[[t]]$, one must be attentive to units. (This is more generally true in linear algebra over a ring R. For instance, if R is an arbitrary

commutative ring with identity, then an element of $GL_n(R)$ is a matrix whose determinant is a unit in R.) In particular, an element $r(t) \in \mathbb{C}[[t]]$ is a unit if and only if the constant term of r is nonzero. [1, Chapter 1, Exercise 5].

Let $e_1, \ldots, e_n$ denote the standard basis of $\mathbb{C}^n$. Given an element $g \in GL_n(\mathbb{C}((t)))$ with entries $g_{ij}(t)$, repeat the following steps to column-reduce g:

- Assume that the strictly upper-triangular entries of the first i rows of g are zero, and that the first i entries along the diagonal of g are $t^{d_1}, t^{d_2}, t^{d_3}, \ldots, t^{d_i}$. (This is reduced form for the first i columns and rows.)
- Consider the entries above and on the diagonal along the $(i+1)$th row of g. Suppose $g_{i+1,j}$ is the entry with the smallest minimal-degree term, and suppose this minimal degree is $t^{d_{i+1}}$. Then $g_{i+1,j}(t) = t^{d_{i+1}}u(t)$, where $u(t)$ is a unit in $\mathbb{C}[[t]]$.

  Replace the jth column g_j of g with the vector $(u(t))^{-1} g_j$.
- For each $i' \in \{i+1, i+2, \ldots, n\}$ other than j the minimal-degree term of $g_{i+1,i'}$ is by definition a multiple of $t^{d_{i+1}} t^{a_{i'}}$ for some $a_{i'} \geq 0$. Hence $g_{i+1,i'} = t^{d_{i+1}} t^{a_{i'}} u'(t)$ for some unit $u'(t) \in \mathbb{C}[[t]]$.

  Replace the (i')th column $g_{i'}$ with $g_{i'} - t^{a_{i'}} u'(t) g_j$.
- **Exchange** columns j and $i+1$.

After this step is implemented, the first $i+1$ columns and rows are in reduced form, so the algorithm can be repeated.

Once the previous algorithm has been repeated n times, we obtain a lower-triangular matrix with entries $t^{d_1}, t^{d_2}, \ldots, t^{d_n}$ along the diagonal, as in Example 3.3.

We may back-eliminate in a similar way, replacing the jth column g_j by an appropriate combination of the columns $g_j, g_{j+1}, \ldots, g_n$ so that each entry $g_{i,j}$ is a Laurent polynomial with *maximal-degree* term $t^{d_i - 1}$.

After this process, the column-reduced form of $g \in GL_n(\mathbb{C}((t)))$ is a matrix

$$\begin{pmatrix} t^{d_1} & 0 & \cdots & 0 \\ g_{2,1} & t^{d_2} & \cdots & 0 \\ & & \cdots & \\ g_{n,1} & g_{n,2} & \cdots & t^{d_n} \end{pmatrix}$$

where $g_{i,j} = 0$ if $i < j$, the entries $g_{i,i} = t^{d_i}$ along the diagonal, and $g_{i,j}$ is a Laurent polynomial with maximal-degree term $t^{d_i - 1}$ if $i > j$.

Example 3.3. *We demonstrate this column reduction. Reordering according to the degrees of the first row transforms*

$$\begin{pmatrix} 0 & t & t^{-1} \\ 1+t & t^{-1} + 2t + \sum_{i \geq 1} t^{2i+1} & t^{-1} \\ 3t + 4t^2 + t^4 + t^5 & 3 + 7t^2 + 2t^3 + 4t^4 + 2t^5 + \sum_{i \geq 3}(3t^{2i} + t^{2i+1}) & 4 + t + t^2 + t^3 \end{pmatrix}$$

to the matrix

$$\begin{pmatrix} t^{-1} & 0 & t \\ t^{-1} & 1+t & t^{-1}+2t+\sum_{i\geq 1} t^{2i+1} \\ 4+t+t^2+t^3 & 3t+4t^2+t^4+t^5 & 3+7t^2+2t^3+4t^4+2t^5+\sum_{i\geq 3}(3t^{2i}+t^{2i+1}) \end{pmatrix}$$

Eliminating using the first entry of the first column gives

$$\begin{pmatrix} t^{-1} & 0 & 0 \\ t^{-1} & 1+t & t^{-1}+\sum_{i\geq 0} t^{2i+1} \\ 4+t+t^2+t^3 & 3t+4t^2+t^4+t^5 & 3+\sum_{i\geq 1}(3t^{2i}+t^{2i+1}) \end{pmatrix}$$

Reordering, rescaling, and eliminating entries in the second row of the last two columns gives

$$\begin{pmatrix} t^{-1} & 0 & 0 \\ t^{-1} & t^{-1} & 1+t \\ 4+t+t^2+t^3 & 3+t^3 & 3t+4t^2+t^4+t^5 \end{pmatrix} \mapsto \begin{pmatrix} t^{-1} & 0 & 0 \\ t^{-1} & t^{-1} & 0 \\ 4+t+t^2+t^3 & 3+t^3 & t^2 \end{pmatrix}$$

In this case, the exponents along the diagonal are $-1, -1, -2$. *Back-eliminating along the last row and then second-to-last row then gives*

$$\begin{pmatrix} t^{-1} & 0 & 0 \\ t^{-1} & t^{-1} & 0 \\ 4+t & 3 & t^2 \end{pmatrix} \mapsto \begin{pmatrix} t^{-1} & 0 & 0 \\ 0 & t^{-1} & 0 \\ 1+t & 3 & t^2 \end{pmatrix}$$

By construction, the reduced form of g is in the coset $gGL_n(\mathbb{C}[[t]])$. In fact, more is true.

Proposition 3.4. *The reduced form of the coset* $gGL_n(\mathbb{C}[[t]])$ *is unique.*

Proof. Suppose that g' and g'' are two different reduced representatives for g, namely that

(a) there are matrices $h', h'' \in GL_n(\mathbb{C}[[t]])$ with $g' = gh'$ and $g'' = gh''$
(b) both g' and g'' are lower-triangular
(c) both g' and g'' have $t^{d_1}, t^{d_2}, \ldots, t^{d_n}$ along the diagonal
(d) for both g' and g'', each lower-triangular entry (i, j) has maximal-degree term $t^{d_i - 1}$

Both g' and g'' are invertible in $GL_n(\mathbb{C}((t)))$. The set of lower-triangular matrices in $GL_n(\mathbb{C}((t)))$ forms a subgroup. In particular, the inverse $(g'')^{-1}$ is also lower-triangular, as is the product $(g'')^{-1}g'$. The diagonal entries of $(g'')^{-1}g'$ agree with those of $(g'')^{-1}g''$ so the matrix $(g'')^{-1}g'$ has ones along the diagonal.

Moreover, the product $(g'')^{-1}g' = (h'')^{-1}h'$ is in $GL_n(\mathbb{C}[[t]])$. In other words, the product $(g'')^{-1}g' = h$, where $h \in GL_n(\mathbb{C}[[t]])$ is a lower-triangular matrix with ones along the diagonal. Writing the product $g' = g''h$, we conclude that the $(j+1, j)$ subdiagonal entry in g' differs from the corresponding entry in g'' by a power series of the form $h_{j+1,j}(t)t^{d_{j+1}}$ for some $h_{j+1,j}(t) \in \mathbb{C}[[t]]$. This violates condition (d)

above, unless $h_{j+1,j}(t) = 0$. Working backwards along subdiagonals, we conclude that actually h is the identity, and $g' = g''$ as desired. □

3.2 Geometric Linear Model for the Affine Grassmannian

We now extend our analogy between ordinary and affine Grassmannians to include the geometric linear model for Grassmannians, as certain subspaces of a vector space. In the ordinary Grassmannian, we obtain a k-dimensional linear subspace from the coset aP by constructing the span of the first k columns of a. Similarly, we can obtain a subspace from the coset $gGL_n(\mathbb{C}[[t]])$ by asking for the $\mathbb{C}[[t]]$-linear span of the columns of g. However, in the ordinary Grassmannian, we have an intrinsic description of the subspaces that appear, namely the set of all k-dimensional linear subspaces of a fixed n-dimensional vector space. The intrinsic description of the subspaces for the affine Grassmannian is more complicated than for ordinary Grassmannians; we turn our attention to this description now.

The next proposition characterizes subspaces in $\mathbb{C}((t)) \otimes \mathbb{C}^n$ that give elements of the affine Grassmannian Gr_n. Billey and Mitchell call this the *Quillen model* for the affine Grassmannian [4, page 210].

For $g \in GL_n(\mathbb{C}((t)))$ with columns $g_1, g_2, \ldots, g_n$, consider the map

$$gGL_n(\mathbb{C}[[t]]) \mapsto L_g = span_{\mathbb{C}[[t]]}\{g_1, g_2, \ldots, g_n\}. \tag{3}$$

The columns of each element of the coset $gGL_n(\mathbb{C}[[t]])$ are invertible $\mathbb{C}[[t]]$-linear combinations of the columns of g so the map $gGL_n(\mathbb{C}[[t]]) \mapsto L_g$ is well-defined.

Proposition 3.5. *The map given in* (3) *is a bijection between the set of cosets* $gGL_n(\mathbb{C}[[t]]) \in GL_n(\mathbb{C}((t)))/GL_n(\mathbb{C}[[t]])$ *and subspaces* $L \subseteq \mathbb{C}((t)) \otimes \mathbb{C}^n$ *satisfying*

(a) L *is a* $\mathbb{C}[t]$*-module, namely* $tL \subseteq L$*, and*
(b) L *differs from* $\mathbb{C}[[t]] \otimes \mathbb{C}^n$ *only in a finite window, namely there exists* $N > 0$ *such that*

$$t^N\left(\mathbb{C}[[t]] \otimes \mathbb{C}^n\right) \subseteq L \subseteq t^{-N}\left(\mathbb{C}[[t]] \otimes \mathbb{C}^n\right).$$

Proof. First we prove that for every g, the subspace L_g satisfies Conditions (a) and (b). Condition (a) holds by construction, since L_g is closed under multiplication by elements of $\mathbb{C}[[t]]$ and hence in particular by t. Without loss of generality, take g to have the canonical form described in Sect. 3.1. Thus the entries of the ith row of g have coefficients in $\mathbb{C}((t))$ with maximum degree t^{d_i-1}. Let d be the minimum exponent that appears in any of the entries $g_{ij}(t)$ of g. Then by definition, each polynomial in each column is in

$$t^d\mathbb{C}[t] = span_{\mathbb{C}}\left\{t^d, t^{d+1}, t^{d+2}, \ldots\right\}.$$

So each g_j is in $t^d(\mathbb{C}[t]) \otimes \mathbb{C}^n$ and hence L_g is in $t^d\mathbb{C}[[t]] \otimes \mathbb{C}^n$.

Similarly, let d' be the maximum exponent so that $t^{d'}$ appears in any of the $g_{ij}(t)$. (Note that d' must be along the diagonal, since each $g_{ij}(t)$ has maximum degree less than the exponent of the diagonal element in its row.) The subspace L_g contains the vectors $t^{d'-d}g_j$ for all $j = 1, 2, \ldots, n$ because $d' - d \geq 0$. Moreover, the minimal-degree term in each $t^{d'-d}g_{ij}(t)$ has degree at least d' which by construction is at least d_j. Thus there is a linear combination of the (lower-triangular) columns $g_1, g_2, \ldots, g_n$ and an exponent d'_j so that

$$t^{d'-d}g_j + p_{j,j+1}g_{j+1} + p_{j,j+2}g_{j+2} + \cdots + p_{j,n}g_n = t^{d'_j}e_j$$

for each j, where $p_{j,j+1}, \ldots, p_{j,n} \in \mathbb{C}[t]$. So L_g contains the elements $t^{d'_1}e_1, t^{d'_2}e_2, \ldots, t^{d'_n}e_n$ for some exponents $d'_1, \ldots, d'_n$. In particular, L_g contains $t^{d''}\mathbb{C}[[t]] \otimes \mathbb{C}^n$ for $d'' = max\{d'_1, d'_2, \ldots, d'_n\}$. If $N = \max\{|d|, d''\}$ then

$$t^N \left(\mathbb{C}[[t]] \otimes \mathbb{C}^n\right) \subseteq L_g \subseteq t^{-N} \left(\mathbb{C}[[t]] \otimes \mathbb{C}^n\right).$$

Next we show that every subspace of the desired form arises as L_g for some matrix g in $GL_n(\mathbb{C}((t)))$. We do this by explicitly constructing the columns $g_1, g_2, \ldots$ of g inductively from L. Suppose that $g_1, \ldots, g_{i-1} \in L$ have been chosen such that every $v \in L$ can be written as

$$v = q_1(t)g_1 + \cdots + q_{j-1}(t)g_{j-1} + r_j(t)e_j + \cdots + r_n(t)e_n \tag{4}$$

where $q_1(t), \ldots, q_{j-1}(t) \in \mathbb{C}[[t]]$ and $r_j(t), \ldots r_n(t) \in \mathbb{C}((t))$.

Choose $v_j \in L$ to be any vector for which the minimum exponent d_j of $r_j(t)$ in Equation (4) is minimal. Since $L \subseteq t^{-N}\mathbb{C}[[t]] \otimes \mathbb{C}^n$, every power of t that appears is at least $-N$ and so this degree d_j exists. By construction, $(r_j(t)t^{-d_j}) \in \mathbb{C}[[t]]$ has nonzero constant term and hence is a unit. Define

$$g_j = (r_j(t)t^{-d_j})^{-1}(v_j - q_1(t)g_1 - \cdots - q_{j-1}(t)g_{j-1}) = t^{d_j}e_j + v'_j$$

where $v'_j \in \mathbb{C}((t)) \otimes span_{\mathbb{C}}\left\{e_{j+1}, \ldots, e_n\right\}$. When $j = 1$, this proves the base case. Otherwise note that since d_j was chosen minimally, by adjusting Eq. (4) by a $\mathbb{C}[[t]]$-multiple of g_j, every $v \in L$ can be written

$$v = q_1(t)g_1 + \cdots + q_{j-1}(t)g_{j-1} + q_j(t)g_j + r_{j+1}(t)e_{j+1} + \cdots + r_n(t)e_n$$

where $q_1(t), \ldots, q_j(t) \in \mathbb{C}[[t]]$ and $r_{j+1}(t), \ldots r_n(t) \in \mathbb{C}((t))$.

Continuing in this manner, we obtain $g_1, \ldots, g_n$ where each g_j has leading term $t^{d_j}e_j$. Moreover, the $\mathbb{C}[[t]]$-span of the $g_1, \ldots, g_n$ is the subspace L. By construction, the vectors $g_1, g_2, \ldots, g_n$ are lower-triangular and so linearly independent over $\mathbb{C}[[t]]$. Thus the matrix g whose columns are the g_j is in $GL_n(\mathbb{C}((t)))$ as desired. □

Remark 3.6. *Note that the subspaces L satisfying Conditions (a) and (b) are in fact generated by Laurent* polynomials *since that holds for the coset representatives in Sect. 3.1. (This can be proven directly by back-substituting at the end of the previous proof, as well.)*

Example 3.7. *Continuing Example 3.3, in this case $d = -1$ and $d' = 2$. We obtain the three column vectors*

$$(t^2, 0, t^3 + t^4)^T, (0, t^2, 3t^3)^T, (0, 0, t^5)^T$$

The third vector is simply $t^5 e_3$ so $d'_3 = 5$ in this case. We eliminate the last entry of $t^3 g_1$ and $t^3 g_2$ using polynomial multiples of the original column vector $(0, 0, t^2)^T$. This gives $t^2 e_1$ and $t^2 e_2$ so $d'_1 = d'_2 = 2$ in this case. (In general, we would need another polynomial linear combination to eliminate the second entry of $t^{d'-d} g_1$.)

3.3 Affine Grassmannian Permutations and Skyline Diagrams

In this subsection, we give the coset description corresponding to an affine Grassmannian permutation and describe its associated subspace using the geometric linear model of Sect. 3.2.

For each i with $1 \leq i \leq n-1$, the affine permutations s_i correspond to the usual permutation matrix in $GL_n(\mathbb{C})$, namely the identity matrix with columns i and $i+1$ exchanged. The affine permutation s_0 can also be represented as a matrix in $GL_n(\mathbb{C}((t)))$ that incorporates both a classical permutation (corresponding to the reflection that exchanges 1 and n) and a rescaling by t. Explicitly, the matrix for s_0 is the monomial matrix with ones along the diagonal in rows $2, 3, \ldots, n-1$, has t in the top right corner, and has t^{-1} in the bottom-left corner.

Given an affine Grassmannian permutation w, we give the coset description for $wGL_n(\mathbb{C}[[t]])$. Since the ordinary permutation matrices of S_n are in $GL_n(\mathbb{C}[[t]])$, we may reorder the columns of w so that nonzero elements lie on the diagonal. This means that $wGL_n(\mathbb{C}[[t]])$ corresponds to a diagonal matrix with $t^{d_1}, t^{d_2}, \ldots, t^{d_n}$ along the diagonal, where $\sum_{i=1}^n d_i = 0$. More specifically, under the map (3), a coset $wGL_n(\mathbb{C}[[t]])$ corresponds to

$$L = span_{\mathbb{C}[[t]]}\{t^{d_1} e_1, t^{d_2} e_2, \ldots, t^{d_n} e_n\}$$

such that $\sum d_i = 0$. This is equivalent to the bijection between affine Grassmannian permutations and the set $\mathcal{S}$ of offset sequences $d = (d_1, d_2, \ldots, d_n)$ from Sect. 3.

The action of the affine Weyl group $\widetilde{S}_n$ on the set $\mathcal{S}$ of offset sequences described in (2) is equivalent to the following action on $\mathbb{C}((t)) \otimes \mathbb{C}^n$ on each generator $t^a e_b$:

- If s_i is a simple transposition with $i = 1, \ldots, n-1$ then $s_i(t^a e_b) = t^a e_{s_i(b)}$ where s_i acts on the integers by exchanging i and $i+1$ and fixing all other integers.

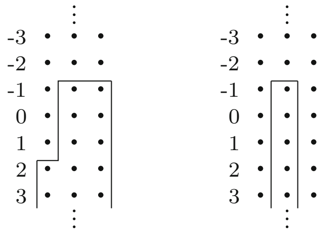
- The simple transposition s_0 acts by

$$s_0(t^a e_1) = t^{a-1} e_n \qquad \text{and} \qquad s_0(t^a e_n) = t^{a+1} e_1 \text{ for each } j$$

and by fixing $t^a e_b$ if $b \neq 1, n$.

One can "sketch" these elements $wGL_n(\mathbb{C}[[t]])$ of the affine Grassmannian using *skyline diagrams*. (David Nadler and Jared Anderson introduced us to this diagrammatic depiction; it also appears in [16, page 59] with the metaphor "above sea level," though the top of our diagram is the bottom of the diagram in [16].) We create a bi-infinite array of dots that describes the generators for $\mathbb{C}((t)) \otimes \mathbb{C}^n$: each dot represents $t^a e_b$ for some a, b, and each row represents $t^a \mathbb{C}^n$ for some a. Figure 1 shows this array for $n = 3$, with rows labeled by the exponent of t.

If $t^a e_b$ is contained in L then Condition (a) of Proposition 3.5 guarantees that all $t^{a'} e_b$ with $a' \geq a$ are in L; in the diagram, this means that once a single dot is in L, so are all dots below and in the same column. Hence we sketch L by outlining the generators that L contains, as shown in Fig. 1a; it is understood that L contains all of the dots below the line. Reading the topmost dots in each column from left to right gives exactly the heights $d_1, \ldots, d_n$.



(a) $L \in \mathrm{Gr}_n$ (b) $L \notin \mathrm{Gr}_n$

Fig. 1 Examples of skyline diagrams

Condition (b) of Proposition 3.5 ensures that every L in Gr_n can be summarized in a finite interval and that every L eventually contains an entire row (and hence, by Condition (1), all the rows beneath). In particular, a skyline diagram like Fig. 1b does NOT represent an element of Gr_n.

Example 3.8. *The example in Fig. 1a represents the plane* $L = span_{\mathbb{C}[[t]]}\{t^2 e_1, t^{-1} e_2, t^{-1} e_3\}$ *and has skyscrapers with heights* $d = (2, -1, -1)$. *This corresponds to the permutation* $s_0 s_2 s_1 s_0$ *in Example 3.1.*

Moreover, the affine Grassmannian can be partitioned according to the exponents of the entries $t^{d_1}, t^{d_2}, \ldots, t^{d_n}$ along the diagonal in the column-reduced form for the general coset $gGL_n(\mathbb{C}[[t]])$. This hints at the decomposition into affine Schubert cells, to which we turn in Sect. 4.

4 Affine Schubert Cells

Similar to the finite case, the affine Grassmannian has a decomposition into affine Schubert cells; these cells can be described similarly to their finite analogues. The most important difference is that many entries in affine Schubert cells for the affine Grassmannian are not free. This has deep implications for our calculations.

We first describe a subgroup of $GL_n(\mathbb{C}[[t]])$ that determines Schubert cells in the affine Grassmannian, analogous to the role of the Borel subgroup B of $GL_n(\mathbb{C})$ in the classical case.

Definition 4.1. *The Iwahori subgroup $I \subseteq GL_n(\mathbb{C}[[t]])$ consists of the elements of $GL_n(\mathbb{C}[[t]])$ that are upper-triangular mod t and invertible mod t.*

Remark 4.2. *The elements of I are precisely those matrices in $GL_n(\mathbb{C}[[t]])$ whose strictly lower-triangular elements are divisible by t, and* none *of whose diagonal entries are divisible by t.*

We give two examples of matrices that are not in I, and one that is:

$$\text{NO: } \begin{pmatrix} 1+t & -t^{-1} \\ t^3 & 1-t \end{pmatrix} \qquad \text{NO: } \begin{pmatrix} t & 1+t \\ t-1 & t \end{pmatrix} \qquad \text{YES: } \begin{pmatrix} 1+t & -1 \\ t^2 & 1-t \end{pmatrix}$$

Definition 4.3. *For an affine Grassmannian permutation w, the* affine Schubert cell *Ω_w is the coset $IwGL_n(\mathbb{C}[[t]])$.*

By Sect. 3.3, we may also write the Schubert cell $IwGL_n(\mathbb{C}[[t]])$ as $It^{\mathbf{d}}GL_n(\mathbb{C}[[t]])$, where $(d_1, \ldots, d_n)$ is the offset sequence $d(w)$ and $t^{\mathbf{d}}$ denotes the diagonal matrix with $t^{d_1}, t^{d_2}, \ldots, t^{d_n}$ along the diagonal.

Using Gaussian elimination, we can describe explicitly the elements of affine Schubert cells. In principle, we would like to conjugate each element g in $It^{\mathbf{d}}$ by an appropriate permutation matrix $\sigma \in S_n$ so that the diagonal entries are reordered in decreasing order. However, while multiplication on the right by σ^{-1} preserves cosets in the affine Grassmannian (because σ is in $GL_n(\mathbb{C}[[t]])$ for each permutation σ), the Iwahori subgroup does not contain σ unless σ is the identity.

Given an affine Grassmannian permutation w whose offset sequence is $d(w) = (d_1, \ldots, d_n)$, let $\mathcal{M}_w$ be the set of matrices in $GL_n(\mathbb{C}((t)))$ that satisfy the following four conditions:

(i) the (j, j) entry is t^{d_j},
(ii) if $d_i < d_j$, the (i, j) entry is zero,
(iii) if $d_i > d_j$, then the (i, j) entry has degree at most $d_i - 1$,
(iv) if the (i, j) entry is nonzero, then its minimum exponent is at least d_j or, if $i > j$, at least $d_j + 1$.

Theorem 4.4. *Let w be an affine Grassmannian permutation whose offset sequence is $(d_1, \ldots, d_n)$. Then elements of the affine Schubert cell Ω_w are in bijection with the*

matrices in $\mathcal{M}_w$ according to the map (3) that sends $g \in \mathcal{M}_w$ to the $\mathbb{C}[[t]]$-span of the columns of g.

Proof. Consider the $\mathbb{C}[[t]]$-span of the columns of a matrix $g \in \mathcal{M}_w$. This corresponds to a coset $gGL_n(\mathbb{C}[[t]])$ by Proposition 3.5. Let $g' = g(t^{\mathbf{d}})^{-1}$. The entries of g' are in $\mathbb{C}[[t]]$ by Condition (iv). The matrix g' is invertible because $\det(g') = \det g \cdot \det(t^{\mathbf{d}})^{-1} = \det g$ is a unit in $\mathbb{C}((t))$ that is also an element of $\mathbb{C}[[t]]$. Thus $g' \in GL_n(\mathbb{C}[[t]])$. Condition (i) implies that g' has ones along the diagonal and Condition (iv) implies that the lower-triangular entries of g' are divisible by t. Therefore g' is in I and so $g \in It^{\mathbf{d}}$. Hence the $\mathbb{C}[[t]]$-span of the columns of a matrix in $\mathcal{M}_w$ is an element of $\Omega_w = It^{\mathbf{d}}GL_n(\mathbb{C}[[t]])$.

In the other direction, for any $\gamma = (\gamma_{ij}(t)) \in I$, consider $\gamma t^{\mathbf{d}}GL_n(\mathbb{C}[[t]])$. We wish to find a matrix $g \in \mathcal{M}_w$ such that $gGL_n(\mathbb{C}[[t]]) = \gamma t^{\mathbf{d}}GL_n(\mathbb{C}[[t]])$.

By Remark 4.2 $\gamma_{jj}(t)$ is a unit in $\mathbb{C}[[t]]$ for all j and $\gamma_{ij}(t) = 0 \mod t$ for $i > j$. By rescaling each column of $\gamma t^{\mathbf{d}}$, i.e., multiplying on the right by the diagonal matrix with entries $(\gamma_{jj}(t))^{-1}$, there is a representative $g^0 \in \gamma t^{\mathbf{d}}GL_n(\mathbb{C}[[t]])$ whose jth column can be written

$$g^0_{1j}(t)e_1 + \cdots + g^0_{j-1,j}(t)e_{j-1} + t^{d_j}e_j + g^0_{j+1,j}(t)e_{j+1} + \cdots + g^0_{nj}(t)e_n \qquad (5)$$

where $g^0_{jj}(t) = t^{d_j}$ and if nonzero, the entries $g^0_{ij}(t) \in \mathbb{C}[[t]]$ have minimal degree at least d_j when $i < j$ and $d_j + 1$ when $i > j$. In other words, the matrix g^0 satisfies Conditions (i) and (iv).

Choose an ordering $i_1, \ldots, i_n$ of $\{1, 2, \ldots, n\}$ so that $d_{i_1} \leq d_{i_2} \leq \cdots \leq d_{i_n}$. We proceed by induction. Given a matrix $g^{k-1} \in \gamma t^{\mathbf{d}}GL_n(\mathbb{C}[[t]])$ that satisfies Condition (i) for all diagonal entries and Conditions (ii)–(iv) for rows $i = i_1, \ldots, i_{k-1}$, we will describe column operations that produce a matrix g^k satisfying Condition (i) for all diagonal entries and Conditions (ii)–(iv) for rows $i = i_1, \ldots, i_{k-1}, i_k$.

For the base case, consider row i_1 of the matrix g^0. By minimality of d_{i_1}, we can eliminate all nondiagonal entries in row i_1 by adding multiples of the i_1th column. In particular, since the jth column looks like (5), we can eliminate the entry $g^0_{i_1,j}(t)$ by scaling as indicated in Fig. 2. In both cases, we add a vector whose entries have the same minimum exponents as the corresponding nondiagonal entry in the jth column, so Condition (iv) remains valid. Also in both cases, the diagonal entry of the jth column is modified by a power series that is divisible by t^{d_j+1} so its minimum entry is still t^{d_j}. Thus if needed, we may rescale the jth column by a unit in $\mathbb{C}[[t]]$ so that t^{d_j} is on the diagonal again. So g^1 satisfies Condition (i) for all diagonal entries and Conditions (ii)–(iv) for row $i = i_1$.

Now suppose that $g^{k-1} \in \gamma t^{\mathbf{d}}GL_n(\mathbb{C}[[t]])$ satisfies Condition (i) for all rows and Conditions (ii)–(iv) for rows $i = i_1, \ldots i_{k-1}$. The i_kth column of g^{k-1} can be written

$$g^{k-1}_{1,i_k}(t)e_1 + \cdots + t^{d_{i_k}}e_{i_k} + \cdots + g^{k-1}_{n,i_k}(t)e_n.$$

The entries of row i_k have the form $g^{k-1}_{i_k,j}(t)$. We eliminate terms of degree at least d_{i_k} in these entries by adding appropriate multiples of the i_kth column of g^{k-1}, using a

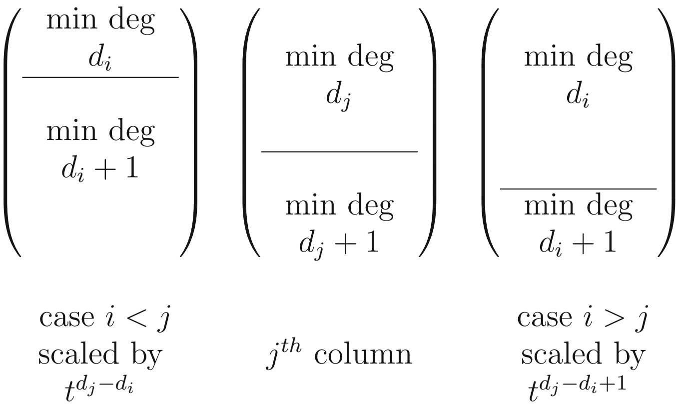


Fig. 2 Schematic for two relative cases of ith and jth columns

polynomial that is divisible by $t^{d_j - d_{i_k}}$ for $i_k < j$ or that is divisible by $t^{d_j + 1 - d_{i_k}}$ for $i_k > j$. By construction, Conditions (i)–(iv) hold for the resulting matrix g^k when we restrict our attention to the i_kth row.

Note that Condition (ii) for $i = i_1, \ldots, i_{k-1}$ for the matrix g^{k-1} implies that the entries $g^{k-1}_{i_1, i_k}(t) = \cdots = g^{k-1}_{i_{k-1}, i_k}(t) = 0$ since $d_{i_1} \leq \cdots \leq d_{i_k - 1} \leq d_{i_k}$. Thus adding $\mathbb{C}[[t]]$-multiples of column i_k leaves the entries in rows $i_1, \ldots, i_{k-1}$ unchanged. Therefore, Conditions (i)–(iv) hold for the matrix g^k for $i = i_1, \ldots, i_{k-1}$ as well.

Finally, by the schematic in Fig. 2 and the same argument as for g^1, the diagonal entry in the jth column for $j > i_k$ can only be changed by terms of degree strictly greater than d_j, so we may rescale the jth column for $j > i_k$ by a unit to preserve Condition (i). Thus Condition (i) holds for the matrix g^k for all rows.

Taking $g = g^n$, this completes the proof. □

In this geometric description of affine Schubert cells, if we set all nondiagonal entries of the matrix g to zero, we obtain the plane

$$L = span_{\mathbb{C}[t]]}\{t^{d_1} e_1, t^{d_2} e_2, \ldots, t^{d_n} e_n\}$$

corresponding to a skyline diagram. This is consistent with the geometric analogy between affine and ordinary Schubert cells: each is described as the span of a collection of vectors parametrized freely by variables; setting those variables to zero gives the unique Weyl group element in the cell.

Example 4.5. *Consider the case when the offset sequence is* $(-1, -1, 2)$ *and* γ *is*

$$\begin{pmatrix} 1 + t & 1 & 0 \\ t & 1 & 0 \\ t + 5t^2 + t^3 & 4t + t^2 & 1 - t \end{pmatrix}$$

The matrix g^0 is the product

$$\begin{pmatrix} t^{-1}+1 & t^{-1} & 0 \\ 1 & t^{-1} & 0 \\ 1+5t+t^2 & 4+t & t^2-t^3 \end{pmatrix} \begin{pmatrix} \sum_{i\geq 0}(-t)^i & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sum_{i\geq 0} t^i \end{pmatrix} = \begin{pmatrix} t^{-1} & t^{-1} & 0 \\ \sum_{i\geq 0}(-t)^i & t^{-1} & 0 \\ (1+5t+t^2)(\sum_{i\geq 0}(-t)^i) & 4+t & t^2 \end{pmatrix}$$

Using the original ordering of the columns, we obtain g^1 by scaling the middle column of the following by $1+t$:

$$\begin{pmatrix} t^{-1} & 0 & 0 \\ \sum_{i\geq 0}(-t)^i & t^{-1}-\sum_{i\geq 0}(-t)^i & 0 \\ (1+5t+t^2)(\sum_{i\geq 0}(-t)^i) & 4+t-(1+5t+t^2)(\sum_{i\geq 0}(-t)^i) & t^2 \end{pmatrix}$$

In other words

$$g^1 = \begin{pmatrix} t^{-1} & 0 & 0 \\ \sum_{i\geq 0}(-t)^i & t^{-1} & 0 \\ (1+5t+t^2)(\sum_{i\geq 0}(-t)^i) & (4+t)(1+t)-(1+5t+t^2) & t^2 \end{pmatrix}$$

In fact, the $(3, 2)$ entry simplifies to 3. To get g^2 we subtract $t\sum_{i\geq 0}(-t)^i$ times the second column from the first:

$$g^2 = \begin{pmatrix} t^{-1} & 0 & 0 \\ 0 & t^{-1} & 0 \\ (1+2t+t^2)(\sum_{i\geq 0}(-t)^i) & 3 & t^2 \end{pmatrix}$$

Finally, we eliminate all terms of degree at least 2 from the third entry of the first column to get g^3:

$$g^3 = \begin{pmatrix} t^{-1} & 0 & 0 \\ 0 & t^{-1} & 0 \\ 1+t & 3 & t^2 \end{pmatrix}$$

In other words, this continues Example 3.3.

5 Descriptions of Affine Schubert Cells: Bit Sequence, Linear Model, Cores, and Partitions

This section defines the combinatorial concepts of the core and the partition associated to an affine Grassmannian permutation w and relates these to the linear algebraic model of the previous section.

Much of the combinatorics of the (co)homology of affine Grassmannians is formulated in terms of cores and partitions. We first describe a bi-infinite bit sequence $p(w)$ associated to an affine Grassmannian permutation w and called the *edge sequence* of

w [16]. The edge sequence $p(w) = \ldots p_{-1}p_0p_1p\ldots$ is given by setting $p_{nd+w(i)} = 0$ for $1 \leq i \leq n$ and $d \geq 0$ and all other $p_j = 1$. We often write $p = \ldots p_{-1}|p_0p_1p\ldots$, inserting the symbol | to form a line between p_{-1} and p_0.

The diagram of the *core* is constructed by tracing from northwest to southeast, with a south step for every one and an east step for every zero in $p(w)$; in this sense, it is like a Young diagram. For instance, the edge sequence $p(id) = \ldots 11111|10000\ldots$ corresponds to the empty diagram. The core can also be constructed using the natural action of $\widetilde{S}_n$ on the bit sequences $p(w)$ induced from the association between $\widetilde{S}_n$ and functions $\mathbb{Z} \to \mathbb{Z}$. In other words, applying the reflections of a reduced word of w from right to left to $p(id)$ yields the bit sequence $p(w)$ as well as its associated core.

Let $\prec$ be the total order on the generators of $\mathbb{C}((t)) \otimes \mathbb{C}^n$

$$\{t^a e_b : a \in \mathbb{Z}, 1 \leq b \leq n\}$$

defined by

$$\cdots \prec t^{a-1}e_n \prec t^a e_1 \prec t^a e_2 \prec \cdots \prec t^a e_n \prec t^{a+1}e_1 \prec \cdots .$$

When we rewrite the results of the previous section in terms of this total order, the Gaussian elimination used is essentially ordinary Gaussian elimination. We will see that the bit sequence $p(w)$ is encoded in the linear algebraic model. Moreover, the reindexing of the generators lets us read the associated core and partition almost immediately. Section 6 uses ideas similar to these in the context of the flag variety.

More precisely, we write the generators $\{t^a e_b\}$ as the totally ordered set $\{f_j\}_{j\in\mathbb{Z}}$ reindexed via the correspondence $f_{na+b} \leftrightarrow t^a e_b$ for all $a \in \mathbb{Z}$ and $1 \leq b \leq n$. This bijection satisfies the following:

- the total order on the generators $\{t^a e_b\}$ of $\mathbb{C}((t)) \otimes \mathbb{C}^n$ is equivalent to the total order on the $\{f_j\}$, in the sense that $t^a e_b \prec t^{a'} e_{b'}$ if and only if $na+b < na'+b'$, and
- the $\mathbb{C}[t]$-module structure on $\mathbb{C}((t)) \otimes \mathbb{C}^n$ gives a $\mathbb{C}[t]$-module structure defined by $t \cdot f_j = f_{j+n}$

With this notation and reinterpreting Conditions (i)–(iv) in terms of the total order $\prec$, we can rewrite Theorem 4.4 as follows. We describe the generators of subspaces using two different bases, though the identification between the bases is so natural as to sometimes make this distinction confusing. We also describe two different generators, the second of which eliminates some unnecessary repetition.

Corollary 5.1. *Let w be an affine Grassmannian permutation w whose offset sequence is $(d_1, \ldots, d_n)$. Then elements of the affine Schubert cell Ω_w are in bijection with the collection of subspaces L that are the $\mathbb{C}[[t]]$-span of*

- *the vectors $v_1, \ldots, v_n$ given by*

$$v_j = t^{d_j}e_j + \sum \beta_{j,a,b}t^a e_b$$

where the sum is over $t^{d_j}e_j \prec t^a e_b$ and $a < d + b$, or

- *the vectors* $v'_1, \ldots, v'_n$ *given by*

$$v'_j = f_{nd_j+j} + \sum_{nd_j+j<k,\, p_k=1} \alpha_{j,k} f_k.$$

Proof. First we confirm that the four conditions for $\mathcal{M}_w$ from Theorem 4.4 are equivalent to the conditions on v_j. The requirements that both $t^{d_j} e_j \prec t^a e_b$ and $a < d_b$ for $\beta_{j,a,b}$ to be nonzero mean that t^{d_j} is the only nonzero coefficient of e_j in v_j. This is Condition (i). Conditions (ii) and (iv) are captured by $t^{d_j} e_j \prec t^a e_b$ and Condition (iii) by $a < d_b$.

We find the $\{v'_j\}$ by rewriting $\{v_j\}$ in terms of the generators $\{f_k\}$ and then adding a convenient normalizing condition. If f_k corresponds to $t^a e_b$ then the condition $t^{d_j} e_j \prec t^a e_b$ corresponds to $nd_j + j < k$. Thus we need only determine a condition equivalent to $a < d_b$. The affine Schubert cell is the $\mathbb{C}[[t]]$-span of the vectors $v_1, \ldots, v_n$ so it is also generated by the larger set of vectors $t^d \cdot v_j$ with $d \geq 0$ and $1 \leq j \leq n$. The leading term of the generator $t^d \cdot v_j$ is $t^{d+d_j} e_j$ and the set of $t^{d+d_j} e_j$ corresponds exactly to those f_k with $p_k = 0$, so we will consider only generators with zero in those entries. This gives the second condition. In other words, write the matrix $(v_1, v_2, \ldots, tv_1, tv_2, \ldots, t^2 v_1, t^2 v_2, \ldots)$ in terms of the generators f_k and then use "column operations" to obtain a reduced echelon form with columns

$$v'_j = f_{nd_j+j} + \sum_{nd_j+j<k,\, p_k=1} \alpha_{j,k} f_k$$

and

$$t^d \cdot v'_j = f_{nd+(nd_j+j)} + \sum_{nd+(nd_j+j)<k,\, p_k=1} \alpha_{j,k-nd} f_k$$

for each j with $1 \leq j \leq n$ and each $d \geq 0$. (Those "column operations" are in fact equivalent to forming a linear combination of the $v_1, v_2, \ldots, v_n$ over $\mathbb{C}[[t]]$ so this process is algebraically valid.) The $\mathbb{C}[[t]]$-span of the $v'_1, v'_2, \ldots, v'_n$ is the same as the $\mathbb{C}[[t]]$-span of the set $v'_1, v'_2, \ldots, tv'_1, tv'_2, \ldots, t^2 v'_1, t^2 v'_2 \ldots$, proving the claim. $\square$

The shape formed by the variables in the *reduced echelon form* for the matrix

$$[v_1, v_2, \ldots, tv_1, tv_2, \ldots]$$

is exactly the core. The independent free variables are precisely $\alpha_{j,k}$ with $1 \leq j \leq n$ and $p_k = 1$; these appear exactly once in $v'_1, v'_2, \ldots, v'_n$. Thus restricting the reduced echelon form for $(v_1, v_2, \ldots, tv_1, tv_2, \ldots)$ to the columns associated to $v_1, v_2, \ldots, v_n$ gives an associated partition, which is clearly $(n-1)$-bounded since no row has free entries in more than $n-1$ columns.

The following examples demonstrate this correspondence. In our examples, we write the reduced vectors only until we reach that particular reduced vector $t^d v'_j$ for

which all generators that come after the pivot of $t^d v'_j$ are contained in the affine Schubert cell. We also use the order $\prec$ on the vectors' leading terms to obtain a lower-triangular matrix; each column sums to one of $v'_1, v'_2, \ldots, tv'_1, tv'_2, \ldots$.

Example 5.2. *For $w = s_0$ with window $[0, 2, 4]$, the affine Schubert cell in $GL_3(\mathbb{C}((t)))$ corresponds to $te_1, e_2, t^{-1}e_3$ and bit sequence $p(w) = \ldots 11|\mathbf{01000}00\ldots$. It is given by the $\mathbb{C}[[t]]$-span of $v_1 = te_1$, $v_2 = e_2$, $v_3 = t^{-1}e_3 + \alpha e_1$, or equivalently by the $\mathbb{C}[[t]]$-span of $v'_1 = f_4$, $v'_2 = f_5$, $v'_3 = f_2 + af_1$, tv'_3. The reduced echelon form for the range $0 \le k \le 4$ (bolded in the above bit sequence) is*

$$\begin{array}{c} t^{-1}e_3 \\ e_1 \\ e_2 \\ e_3 \\ te_1 \end{array} \begin{pmatrix} 1 & 0 & 0 & 0 \\ \alpha & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{array}{c} f_0 \\ f_1 \\ f_2 \\ f_3 \\ f_4 \end{array}$$

where the columns give the generators $v_3, v_2, tv_3 - \alpha v_1, v_1$, respectively. (We think of the third column as associated to tv_3 even though it is in fact a linear combination with leading term the same as that of tv_3.) In this case, the core and partition coincide and are both (1), representing the unique (and free) variable.

Example 5.3. *For the window $w = [-1, 0, 7]$, $p_{-1} = p_2 = p_5 = \cdots = 0$, $p_0 = p_3 = p_6 = \cdots = 0$, and $p_7 = p_{10} = \cdots = 0$, so that $p(w) = \ldots 1111\mathbf{0}|\mathbf{0100100}00 \ldots$. The reduced echelon form has all zeros for $k < -1$, and the range $-1 \le k \le 7$ whose p-values are in bold above, is pictured below.*

$$\begin{array}{c} \\ t^{-1}e_2 \\ t^{-1}e_3 \\ e_1 \\ e_2 \\ e_3 \\ te_1 \\ te_2 \\ te_3 \\ t^2e_1 \end{array} \begin{array}{c} \begin{array}{ccccccc} v'_2 & v'_3 & tv'_2 & tv'_3 & t^2v'_2 & t^2v'_3 & v'_1 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{1,1} & \alpha_{2,1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \alpha_{1,4} & \alpha_{2,4} & \alpha_{1,1} & \alpha_{2,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{array}{c} \\ f_{-1} \\ f_0 \\ f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \end{array} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ * & * & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ * & * & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The core $(4, 2)$ comes from all of the nonzero variables in this cell. The free variables are indexed by restricting our attention to the columns corresponding to v'_2, v'_3 and v'_1 as sketched in the matrix on the right. In this case, the associated bounded partition is $(2, 2)$. (Examples 3.3 and 4.5 both concerned an element in this cell.)

Example 5.4. *For the window $w = [-4, 3, 7]$, $p_{-4} = p_{-1} = p_2 = p_5 = \cdots = 0$, $p_3 = p_6 = \cdots = 0$, and $p_7 = p_{10} = \cdots = 0$, so that $p(w) = \ldots 11110110|11001000000 \ldots$. The associated core is $(4, 2, 2, 1, 1)$.*

In sum, affine Grassmannians have affine Schubert cells that can be described linear algebraically by analogy with the Schubert cells in the finite Grassmannian. The finite Schubert cells are indexed by either of two sets of combinatorial objects: n-bit sequences with zeros (corresponding to the basis vectors contained in the Schubert cell) or Young diagrams (corresponding to the shape created by the free entries in the Schubert cell). Similarly, the affine Schubert cells are indexed by either: sequences $(d_1, \ldots, d_n) \in \mathbb{Z}^n$ with $\sum d_j = 0$ (corresponding to the exponents along the diagonal of the affine Grassmannian permutation) or Young diagrams (corresponding to the entries in the Schubert cell parametrized by variables). Here, the dimension of an affine Schubert cell Ω_w is equal to the length $\ell(w)$, which is also equal to the number of boxes in the associated bounded partition.

6 (Grassmannian) Hessenberg Schubert Cells and Varieties

In this section, we review several key facts and definitions about Hessenberg varieties. We use notation and terminology from type A, though many of these facts extend to general Lie type. We define Hessenberg Schubert cells in the flag variety and their images when projected to a Grassmannian, which we call Grassmannian Hessenberg Schubert cells. We show that a class of Hessenberg Schubert cells are isomorphic to affine space, a fact that we use in Sect. 7. Finally we give conditions for the image of the closure of a Hessenberg Schubert cell to be the closure of the corresponding Grassmannian Hessenberg Schubert cell.

6.1 Hessenberg Varieties

Let $V_\bullet = V_1 \subseteq V_2 \subseteq \cdots \subseteq V_n = \mathbb{C}^n$ be a complete flag in $\mathbb{C}^n$. The space of such flags forms the complete flag variety $Fl(n)$ which can also be described as $GL_n(\mathbb{C})/B$ where B is the Borel subgroup of invertible $n \times n$ upper-triangular matrices.

In this and the next section, we use the upper-triangular Borel to describe the flag variety. This leads to different conventions for combinatorics and linear algebra than in the previous sections. Reconciling these notations is the key point of the results in the last section.

Definition 6.1. *Fix a linear operator $X : \mathbb{C}^n \to \mathbb{C}^n$. Fix a nondecreasing function $h : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ with $h(i) \geq i$ for all i. The* Hessenberg variety of X and h *is denoted $\mathcal{H}(X, h)$ and defined to be*

$$\mathcal{H}(X, h) = \{V_\bullet : XV_i \subseteq V_{h(i)}\}.$$

Here, the function h is called a Hessenberg function.

Equivalently, we can describe Hessenberg varieties using Lie theory. Write M_n for the set of $n \times n$ complex matrices and $\mathfrak{b}$ for the set of upper-triangular matrices in M_n. A *Hessenberg space* H is a linear subspace of M_n that satisfies the following two properties:

(1) H contains the upper-triangular matrices $\mathfrak{b}$, and
(2) H is closed under Lie bracket with the upper-triangular matrices, so $[\mathfrak{b}, H] \subseteq H$.

Since Hessenberg spaces are closed under Lie bracket with the diagonal matrices, each Hessenberg space is a sum of root spaces. Hence, each Hessenberg space can be denoted by the subset of roots $\mathcal{R}_H$ satisfying

$$H = \langle E_{ii} : i = 1, 2, \ldots, n \rangle \oplus \bigoplus_{(i,j) \in \mathcal{R}_H} \langle E_{ij} \rangle$$

where E_{ij} denotes the matrix with 1 in the (i, j) entry and zero elsewhere.

Given a linear operator $X : \mathbb{C}^n \to \mathbb{C}^n$ and a Hessenberg space H, the Hessenberg variety of X and H is

$$\mathcal{H}(X, H) = \{mB \in GL_n(\mathbb{C})/B : m^{-1}Xm \in H\}.$$

(Since H is closed under Lie bracket with $\mathfrak{b}$ it is also closed under conjugation by elements of B, so Hessenberg varieties are in fact well-defined.) The two definitions of Hessenberg varieties are equivalent because there is a natural bijection between Hessenberg functions and Hessenberg spaces that is given by

$$H = \langle E_{ij} : i \leq h(j) \text{ for all } j = 1, 2, \ldots, n \rangle.$$

Hessenberg varieties have a natural partial order induced by inclusion of vector spaces in M_n. This is equivalent to the partial order on Hessenberg functions given by

$$h \geq h' \text{ if and only if } h(i) \geq h'(i) \tag{6}$$

for all $i = 1, 2, \ldots, n$.

In this paper, we assume X is nilpotent. In this case, the Jordan canonical form of X determines a Young diagram by the rule that if the Jordan form has blocks of dimension $d_1 \geq d_2 \geq \cdots$ then the Young diagram has rows $d_1 \geq d_2 \geq \cdots$. Our convention is to use Young diagrams that are left-aligned and top-aligned. We now define a particular permutation using the Jordan form of X (Fig. 3).

Definition 6.2. *Fill the boxes in the Young diagram for X with the numbers $\{1, 2, \ldots, n\}$ without repetition, starting at the lower left corner and moving up the first column, then up the second column, and so on. The permuted Jordan form of X is the matrix $X = \sum E_{ij}$ where the sum is taken over pairs $\boxed{i}\boxed{j}$ in the filled Young diagram.*

$$\begin{array}{|c|c|c|}\hline 3 & 5 & 6 \\ \hline 2 & 4 & \multicolumn{1}{c}{} \\ \cline{1-2} 1 & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \qquad \Leftrightarrow \qquad X = E_{2,4} + E_{3,5} + E_{5,6}$$

Fig. 3 An example of permuted Jordan form

6.2 *Hessenberg Schubert Cells and Varieties*

Hessenberg Schubert cells are obtained by intersecting a Hessenberg variety with a Schubert cell in $Fl(n)$. Our convention is to associate a permutation $\sigma \in S_n$ with the permutation matrix (also denoted σ) so that for each i, the ith row of the matrix σ is the basis vector $e_{\sigma(i)}$. The permutation matrix for σ^{-1} is transpose to that for σ so that the permutation σ^{-1} can be read from the columns of σ.

The complete flag variety $Fl(n)$ has a decomposition

$$Fl(n) = \bigcup B\sigma B/B$$

into a disjoint union of *Schubert cells* $C_\sigma := B\sigma B/B$ indexed by $\sigma \in S_n$.

Definition 6.3. *Let $\mathcal{H}(X, h) \subseteq Fl(n)$ be a Hessenberg variety. The* Hessenberg Schubert cell *for $\mathcal{H}(X, h)$ and $\sigma \in S_n$ is the intersection $C_{\sigma,\mathcal{H}(X,h)} := \mathcal{H}(X, h) \cap C_\sigma$. The corresponding* Hessenberg Schubert variety *is the closure $\overline{C_{\sigma,\mathcal{H}(X,h)}}$ in $\mathcal{H}(X, h)$. When X and h are understood, we write $C_{\sigma,\mathcal{H}}$ and $\overline{C_{\sigma,\mathcal{H}}}$.*

A Hessenberg Schubert cell can be a complicated topological space. For instance, Kostant and Rietsch studied an important paving of Peterson varieties created from nonaffine Hessenberg Schubert cells [13, 22]. For our purposes, we want Hessenberg Schubert cells to be affine cells. This is true under the conventions of Proposition 6.4, when X is a nilpotent matrix in permuted Jordan form and the Schubert cells are written with respect to the upper-triangular Schubert decomposition.

The next proposition shows that when X is in permuted Jordan form, each Hessenberg Schubert cell in the flag variety is isomorphic to an affine space and constructs that isomorphism. The proof is not new; the maps identified by the second author in earlier work [24, Lemma 5.2 and Theorem 6.1] are actually isomorphisms, though the earlier paper only used the fact that the maps were homeomorphisms.

Proposition 6.4. *Suppose $\mathcal{H}(X, h)$ is a Hessenberg variety for a nilpotent operator X in permuted Jordan form and a Hessenberg function h. For $\sigma \in S_n$, the Hessenberg Schubert cell $C_{\sigma,\mathcal{H}}$ is nonempty if and only if the flag $\sigma B \in \mathcal{H}(X, h)$, and in this case, $C_{\sigma,\mathcal{H}}$ is isomorphic to an affine space $\mathbb{C}^\ell$ for some $\ell \geq 0$.*

Proof. The fact that the intersection $\mathcal{H}(X, h) \cap C_\sigma$ is nonempty if and only if the flag $\sigma B \in \mathcal{H}(X, h)$ was proved in the original result [24, Theorem 6.1].

We sketch the argument of the isomorphism to affine space, referring to details in [24] as needed, and checking at each stage that the homeomorphisms in that work are in fact isomorphisms.

In [24, Theorem 6.1], Tymoczko showed that the Hessenberg Schubert cell $C_{\sigma,\mathcal{H}}$ can be described as an iterated tower of affine fiber bundles

$$Z_1 \xrightarrow{\pi_1} Z_2 \xrightarrow{\pi_2} Z_3 \xrightarrow{\pi_3} \cdots \xrightarrow{\pi_{n-2}} Z_{n-1}$$

where Z_1 is homeomorphic to the Hessenberg Schubert cell itself, the base space Z_{n-1} is homeomorphic to affine space $\mathbb{C}^{\ell_{n-1}}$ and each map $\pi_i : Z_i \longrightarrow Z_{i+1}$ is a fiber bundle whose fiber is an affine space $\mathbb{C}^{\ell_i}$. Explicitly, each Z_i is defined as a subset of the unipotent group U of upper-triangular matrices with ones along the diagonal, as follows. The group U can be factored uniquely as $U = U_{n-1}U_{n-2}\cdots U_1$ where each U_i is the subset of U whose nondiagonal entries all lie on the ith row. The maps $\pi_i : Z_i \longrightarrow Z_{i+1}$ send the element $u_{n-1}u_{n-2}\cdots u_{i+1}u_i \in Z_i$ to the element $u_{n-1}u_{n-2}\cdots u_{i+1} \in Z_{i+1}$.

In other words, each Z_i is isomorphic (as an algebraic *variety* not a group) to a product space, and each map π_i is the algebraic map given by projection to one factor.

Reference [24, Theorem 6.1] also defined the homeomorphism between Z_1 and the Hessenberg Schubert cell $C_{\sigma,\mathcal{H}}$ as multiplication $\tau \mapsto \tau\sigma$ composed with the quotient map $G \to G/B$ that sends the matrix $\tau\sigma$ to the flag $\tau\sigma B$.

These maps are both algebraic. When we restrict the inverse map to a single Schubert cell, the inverse map from G/B to G is also algebraic. So in fact Z_1 is isomorphic as a variety to $C_{\sigma,\mathcal{H}}$.

Reference [24, Lemma 5.2] describes the fiber bundle structure of each Z_i as follows. Given $u' \in Z_{i+1}$ the fiber $\pi_i^{-1}(u')$ is the set of solutions $\mathbf{x}_{u'}$ to a system of equations $\mathbf{x}_{u'}M_{u'} = \mathbf{v}_{u'}$. The system is affine linear, meaning that each equation is polynomial of degree one and may have a constant term. Both $M_{u'}$ and $\mathbf{v}_{u'}$ vary continuously in u' by conjugation. Reference [24, Lemma 5.2] identified a fixed set of indices I so that the entries in positions I of the matrices in $\pi_i^{-1}(u')$ are free and showed that the map sending $u'u \mapsto (u', (u_{ik})_{k\in I})$ is a homeomorphism with inverse determined by the system $\mathbf{x}_{u'}M_{u'} = \mathbf{v}_{u'}$.

This map is algebraic because it realizes the varieties as product spaces and then projects to certain factors. The inverse map is also algebraic because it is the solution to an affine linear system. Thus each Z_i is isomorphic as a variety to the product of Z_{i+1} with the affine fiber $\mathbb{C}^{\ell_i}$.

Finally the space Z_{n-1} is affine itself because it is a subgroup of U_{n-1} in which certain entries, determined by σ, are set to zero [24, Definition 2.3 and Theorem 6.1]. This shows that the Hessenberg Schubert cell is in fact isomorphic to affine space, of the same dimension as in the original result. □

6.3 Grassmannian Hessenberg Schubert Cells

For a fixed Hessenberg variety $\mathcal{H}(X, h) \subseteq Fl(n)$ and an integer $1 \leq k \leq n$, we describe Grassmannian Hessenberg Schubert cells as projections of Hessenberg Schubert cells to the Grassmannian $Gr(k, n)$.

There is a natural projection map from the complete flag variety to the Grassmannian given by $\pi_k : Fl(n) \to Gr(k, n)$ that sends the flag $V_\bullet$ to its k-dimensional part V_k. If $Gr(k, n)$ is realized as a quotient G/P_k by a maximal parabolic P_k then the projection may be written as the quotient map $\pi_k : G/B \to G/P_k$.

Definition 6.5. *Given k and a Hessenberg Schubert cell $C_{\sigma,\mathcal{H}}$ the* Grassmannian Hessenberg Schubert cell *is the image*

$$\pi_k \left(C_{\sigma,\mathcal{H}}\right)$$

in $Gr(k, n)$. The corresponding Grassmannian Hessenberg Schubert variety *is the closure $\overline{\pi_k \left(C_{\sigma,\mathcal{H}}\right)}$ in $Gr(k, n)$.*

The next lemma follows directly from the corresponding fact for flag varieties.

Lemma 6.6. *The image of the projection $\pi_k : C_{\sigma,\mathcal{H}} \to Gr(k, n)$ is nonempty if and only if the collection $\{V_k : V_\bullet \in C_{\sigma,\mathcal{H}}\}$ contains the k-plane $\langle e_{\sigma^{-1}(1)}, e_{\sigma^{-1}(2)}, \ldots, e_{\sigma^{-1}(k)}\rangle$.*

Proof. Proposition 6.4 stated that $C_{\sigma,\mathcal{H}}$ is nonempty if and only if $\sigma B \in \mathcal{H}(X, h)$. If B consists of the upper-triangular matrices, then the k-dimensional part of the flag σB is $\langle e_{\sigma^{-1}(1)}, e_{\sigma^{-1}(2)}, \ldots, e_{\sigma^{-1}(k)}\rangle$ by definition. The set

$$\{V_k : V_\bullet \in C_{\sigma,\mathcal{H}}\}$$

is the image of $C_{\sigma,\mathcal{H}}$ under the projection to $Gr(k, n)$, so the claim follows. □

6.4 Hessenberg Schubert Varieties and Grassmannian Hessenberg Schubert Varieties

We now give a dimension condition for the image of a Hessenberg Schubert variety under the projection map to be a Grassmannian Hessenberg Schubert variety. More precisely we have the following.

Proposition 6.7. *Let $\mathcal{H} := \mathcal{H}(X, h)$ be a Hessenberg variety in $Fl(n)$ for a nilpotent X in permuted Jordan form and a Hessenberg function h. For $1 \leq k \leq n$ and $\sigma \in S_n$, if the dimension of the Hessenberg Schubert cell $C_{\sigma,\mathcal{H}}$ is equal to the dimension of the Grassmannian Hessenberg Schubert cell $\pi_k(C_{\sigma,\mathcal{H}})$ then*

$$\overline{\pi_k(C_{\sigma,\mathcal{H}})} = \pi_k(\overline{C_{\sigma,\mathcal{H}}}).$$

Proof. The image of a closed irreducible topological space under a closed continuous map is irreducible (and closed). The closed set $\overline{C_{\sigma,\mathcal{H}}}$ is irreducible because $C_{\sigma,\mathcal{H}}$ is isomorphic to affine space, as proven in Proposition 6.4. Thus the image $\pi_k(\overline{C_{\sigma,\mathcal{H}}})$ is a closed irreducible set, and since it contains the Grassmannian Hessenberg Schubert cell $\pi_k(C_{\sigma,\mathcal{H}})$,

$$\overline{\pi_k(C_{\sigma,\mathcal{H}})} \subseteq \pi_k(\overline{C_{\sigma,\mathcal{H}}}). \tag{7}$$

This gives $\dim \overline{\pi_k(C_{\sigma,\mathcal{H}})} \leq \dim \pi_k(\overline{C_{\sigma,\mathcal{H}}}) \leq \dim \overline{C_{\sigma,\mathcal{H}}}$. The dimension of each open set is the same as the dimension of its closure, so by our hypothesis on the dimensions, these inequalities are all in fact equalities. Since $\pi_k(\overline{C_{\sigma,\mathcal{H}}})$ is closed and irreducible, Equation (7) becomes an equality as needed. □

The dimension of a Hessenberg Schubert cell is usually larger than that of a corresponding Grassmannian Hessenberg Schubert cell. However, there are many natural examples when they are the same, e.g., Sect. 7. The previous proposition suggests that when equality holds, the Schubert calculus of Hessenberg varieties could be computed through Schubert calculus in the Grassmannian. Results on the dimension of Hessenberg Schubert cells can be found in [21, 26].

6.5 Minimal Hessenberg Schubert Cells

Before applying the results of Sect. 6 to affine Schubert cells in the next section, we consider *minimal* Hessenberg Schubert cells for fixed X and w and prove that they are well-defined and unique when X is nilpotent. Recall the partial order on Hessenberg functions $h' \leq h$ given by (6).

Definition 6.8. *Fix an $n \times n$ matrix X and a permutation $\sigma \in S_n$. A Hessenberg function h is a* minimal Hessenberg function *for X and σ if*

- *$C_{\sigma,\mathcal{H}(X,h)} \neq \emptyset$ in $Fl(n)$ and*
- *for every Hessenberg function $h' \leq h$ the Hessenberg Schubert cell $C_{\sigma,\mathcal{H}(X,h')}$ is empty.*

In this case we call $C_{\sigma,\mathcal{H}(X,h)}$ a minimal Hessenberg Schubert cell *for X and σ.*

The next result shows that minimal Hessenberg Schubert cells are well-defined and unique.

Proposition 6.9. *Let X be a nilpotent linear operator in permuted Jordan form. The minimal Hessenberg Schubert cell for X and σ is well-defined. In particular, there is a* unique *minimal Hessenberg function h with $C_{\sigma,\mathcal{H}(X,h)} \neq \emptyset$.*

Proof. From Proposition 6.4, the Hessenberg Schubert cell $C_{\sigma,\mathcal{H}(X,h)}$ is nonempty if and only if the flag $\sigma B \in \mathcal{H}(X,h)$. The flag $\sigma B \in \mathcal{H}(X,h)$ if and only if $\sigma^{-1}X\sigma \in H$ by definition.

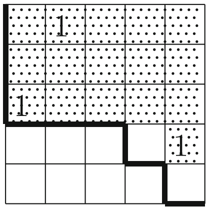


Fig. 4 An example of shadows for $\sigma = (14)(325)$ and $X = E_{13} + E_{24} + E_{45}$ (shaded), with minimal Hessenberg function $h = 33345$ (Hessenberg space outlined)

Each Hessenberg space H is a direct sum of root spaces because it is closed under the adjoint action of $\mathfrak{t}$. Hence if a Hessenberg space H contains an element $\sum c_{ij} E_{ij}$ then it contains every root space $\langle E_{ij} \rangle$ for which $c_{ij} \neq 0$. If $\sigma^{-1} X \sigma \in H$ then H contains $\langle E_{\sigma(i),\sigma(j)} \rangle$ for each pair i, j that appears in the sum $X = \sum E_{ij}$.

In particular, define a Hessenberg space by

$$H = \mathfrak{b} + \bigoplus \langle E_{ij} : i \leq \sigma(i') \text{ and } j \geq \sigma(j') \rangle$$

where the sum is over all pairs $(\sigma(i'), \sigma(j'))$ that appear in $\sigma^{-1} X \sigma = \sum E_{\sigma(i),\sigma(j)}$. This is a Hessenberg space by construction. It contains $\sigma^{-1} X \sigma$ also by construction. Conversely, any Hessenberg space containing $\sigma^{-1} X \sigma$ must contain H.

The Hessenberg function corresponding to this Hessenberg space H is defined by

$$h(i) = \max \begin{cases} i \\ h(i-1) \\ \sigma(j_1) \text{ where } E_{j_1,j_2} \text{ is in permuted Jordan form for } X \text{ and } j_2 = \sigma^{-1}(i). \end{cases}$$

□

Recall that E_{j_1,j_2} is in the permuted Jordan form for X if and only if $\boxed{j_1 \mid j_2}$ is in the Young tableau describing permuted Jordan form (Fig. 4).

The minimal H is easy to describe schematically. If $E_{i',j'}$ is a root vector, define the *shadow* of $E_{i',j'}$ to be the span $\langle E_{ij} : i' \leq i \text{ and } j' \geq j \rangle$. Given an element $X \in M_n$ with $X = \sum c_{ij} E_{ij}$, define the shadow of X to be the sum of the shadows of E_{ij} for which $c_{ij} \neq 0$. Given X and σ, denote the shadow of $\sigma^{-1} X \sigma$ by V. Then the minimal Hessenberg Schubert cell for X and σ is defined using the Hessenberg space $H = V + \mathfrak{b}$.

7 Affine Schubert Cells and Hessenberg Schubert Cells

In this section, we show that every affine Schubert cell Ω_w is isomorphic to a Hessenberg Schubert cell. Denote by h_{id} the Hessenberg function defined by $h_{id}(i) = i$ for all i.

Theorem 7.1. *Every affine Schubert cell is isomorphic to a Hessenberg Schubert cell. More precisely, given an affine Grassmannian permutation w, the corresponding affine Schubert cell Ω_w is isomorphic to a Hessenberg Schubert cell $C_{\sigma,\mathcal{H}(X,h_{id})}$ for some nilpotent linear operator $X : \mathbb{C}^N \to \mathbb{C}^N$ and permutation $\sigma \in S_N$.*

In order to prove the theorem, we first give two constructions identifying Ω_w with a Grassmannian Hessenberg Schubert cell in a finite Grassmannian.

Proposition 7.2. *Given an affine Grassmannian permutation w with corresponding affine Schubert cell Ω_w, there are integers $1 \le K \le N$, a projection map $\pi : Gr_n \to Gr(K, N)$, and a nilpotent linear operator $X : \mathbb{C}^N \to \mathbb{C}^N$ so that:*

(a) there is a Hessenberg function h with corresponding Hessenberg variety $\mathcal{H}_h := \mathcal{H}(X, h)$ and a permutation $\tau \in S_N$ for which the Grassmannian Hessenberg Schubert cell is the projection of the affine Schubert cell to $Gr(K, N)$:

$$\pi_K(C_{\tau,\mathcal{H}_h}) = \pi(\Omega_w)$$

(b) there is a Grassmannian permutation $\overline{\tau} \in S_N$ so that $\overline{\tau}B \in \mathcal{H}_{id} := \mathcal{H}(X, h_{id})$ and

$$\pi_K(C_{\overline{\tau},\mathcal{H}_{id}}) = \pi(\Omega_w)$$

Proof. Given an affine Grassmannian permutation w and its corresponding affine Schubert cell Ω_w we start by defining N, K, and X.

- A positive integer N: Let $t^{a'}e_{b'}$ be the lowest-ordered generator with respect to $\prec$ that is not contained in every plane in Ω_w. Let $t^a e_b$ be the highest-indexed generator that appears as a pivot in Ω_w. Define a'' to be a' if $b - 1 \ge b'$ and $a' + 1$ otherwise. Let $N = n(a'' - a)$ and consider the vector space $\mathcal{N}$ spanned by the N vectors

$$\underbrace{\langle t^a e_b, t^a e_{b+1}, t^a e_{b+2}, \ldots, t^a e_n, t^{a+1}e_1, t^{a+1}e_2, t^{a+1}e_3, \ldots, t^{a+1}e_n, t^{a+2}e_1, \ldots, t^{a'+1}e_{b'} \ldots, t^{a''}e_{b-1}\rangle}_{N \text{ vectors}}.$$

  The vector space $\mathcal{N}$ of dimension N is called the *slice* of the affine Grassmannian. (In principle, we want the slice to be the $\mathbb{C}$-span of the vectors from $t^a e_b$ to $t^{a'+1}e_{b'}$ inclusive, so that each vector $e_1, e_2, \ldots, e_n$ is represented in the slice. However, we expand the slice for notational convenience so that N is divisible by n.)
- A positive integer $K \le N$: Write the vectors spanning a generic plane in the affine Grassmannian cell as

$$\begin{array}{c} v_1, tv_1, \ldots, t^{a_1}v_1 \\ v_2, tv_2, \ldots, t^{a_2}v_2 \\ \cdots \\ v_n, tv_n, \ldots, t^{a_n}v_n \end{array}$$

  where each $a_i \ge 0$ is defined by the condition that the leading term of $t^{a_i}v_i$ is in the slice but $t^{a_i+1}v_i$ is not. Assume without loss of generality that the vectors are indexed satisfying the constraint that

$$\text{leading term of } v_1 \succ \text{ leading term of } v_2 \succ \cdots \succ \text{ leading term of } v_n.$$

Let K be the number of nonzero vectors in this set.

- We identify each vector in the affine Grassmannian with its image under the projection $\pi : \mathrm{Gr}_n \to Gr(K, N)$ inside $\mathcal{N}$. We conflate terminology slightly and say that this is the image of Gr_n inside the *slice*.
- We reorder the nonzero vectors $v_1, tv_1, \ldots, t^{a_1}v_1, v_2, tv_2, \ldots, t^{a_2}v_2, \ldots, v_n, tv_n, \ldots, t^{a_n}v_n$ in decreasing order of leading terms and call these reordered vectors $u_1, u_2, \ldots, u_K$. (Note that $u_1 = v_1$ by assumption that the leading term of v_1 is maximal.)
- A nilpotent linear operator X: Let X be the matrix in permuted Jordan form corresponding to the rectangular $n \times (a'' - a)$ Young diagram, as in Definition 6.2. (By construction, $N = n(a'' - a)$ so the matrix can act on a slice.) Define a bijection between $\mathbb{C}^N$ and the slice under consideration by labeling the basis vectors of $\mathbb{C}^N$ with $t^a e_b$ in the *bottom-left* corner of the Young diagram for X and then in decreasing order up the leftmost column, then up the second-to-left column, etc. Since X is written in permuted Jordan form with respect to this basis, we have $X = \sum_{i=1}^{N-n} E_{i,i+n}$.
  With these conventions, the action of X coincides with the action of t in the following sense:
  - $tu_i = u_j$ if and only if $Xu_i = u_j$
  - tu_i is outside the window if and only if $Xu_i = 0$.

We can now define the matrix of the permutation $\tau \in S_N$. For each integer i with $1 \le i \le K$ the ith column of τ is one in the entry corresponding to the pivot of u_i. Now append the last $N - K$ basis vectors as columns with no descents. Though there are no descents in the last $N - K$ entries, τ is not a Grassmannian permutation: each of the first K entries is a descent by definition of the u_i.

For each $i \in \{1, 2, \ldots, K\}$ define the Hessenberg function h as follows:

- If $tu_i = u_j$ then $h(i) = j$ and
- if tu_i is outside the slice then $h(i) = K$.

For each $i \in \{K + 1, \ldots, N\}$ define $h(i) = i$.

In the first K columns, the conditions used to define h are the same as those used to construct X and in the last $N - K$ columns, the conditions are moot. So $\tau B \in \mathcal{H}(X, h)$.

We now prove that the Grassmannian Hessenberg Schubert cell defined by these parameters is the same as the projection of the affine Schubert cell to the specified slice. Corollary 5.1 says that the affine Schubert cell conditions are equivalent to choosing the entries of $v_1, v_2, \ldots, v_n$ freely subject to the constraint that the leading entry of each vector is in the position specified by τ, and generating the rest of the K vectors by appropriate powers $t^i v_j$. By contrast, the first K columns of the flags in $C_{\tau,\mathcal{H}_h}$ are the same freely chosen $v_1, v_2, \ldots, v_n$ and powers $X^i v_j$ ordered consistent with $u_1, u_2, \ldots, u_K$, while the last $N - K$ columns of the flag are the basis vectors indicated by the last $N - K$ entries of τ. (As before, such a flag is in

$\mathcal{H}_h$ by construction of X.) No other flag that satisfies the Hessenberg conditions also satisfies the affine Schubert cell conditions because for each $i \le K$ adding a nonzero linear combination of the first $i-1$ columns of the flag to the ith column would change the pivot in the ith column, thus moving the flag outside of the cell C_τ.

It follows that $\pi_K(C_{\tau,\mathcal{H}_h})$ is the projection of the affine Schubert cell Ω_w to the specified slice, as desired.

This Hessenberg Schubert cell is "natural" in the sense that it doesn't require any normalization to see that it is the projection of the corresponding affine Schubert cell. To prove part (b), we modify the cell so that the permutation flag corresponds to a Grassmannian permutation. (The vectors in part (b) are the v_1', v_2', v_3', …, though this fact is not explicitly needed in the proof.)

Define $\overline{\tau}$ to be the Grassmannian permutation obtained from the permutation τ by reordering the first K columns so there are no descents.

We now confirm both that $\overline{\tau}B \in \mathcal{H}(X, h_{id})$ and that $\pi_K(C_{\overline{\tau},\mathcal{H}_{id}})$ are the projection $\pi(\Omega_w)$ of the affine Schubert cell Ω_w to the slice under consideration.

First note that $\overline{\tau}B \in \mathcal{H}(X, h_{id})$ because by construction, the basis vectors within each Jordan block of X appear in the flag $\overline{\tau}B$ in increasing order (though not necessarily successively). Thus the image of each of the first i columns of $\overline{\tau}$ under X is either zero or one of the first $i-1$ columns of $\overline{\tau}$.

Now we show that $\pi_K(C_{\overline{\tau},\mathcal{H}_{id}})$ is the projection of Ω_w to the slice under consideration. Let $c_{i_1}, c_{i_2}, \ldots, c_{i_n}$ be columns satisfying the following:

- they are within the first K columns of the matrix of the flag
- there is exactly one column corresponding to a multiple of each of $e_1, e_2, \ldots, e_n$ in the generators $\{t^a e_b\}$ for the affine Grassmannian
- the pivot for each column corresponds to the multiple of $e_1, e_2, \ldots, e_n$ with the lowest power of t within the first K columns of the matrix. (More technically: if the column c_{i_j} corresponds to $t^a e_b$ then none of the first K columns of the matrix have a pivot in the position corresponding to $t^{a-1}e_b$.)

For each i_j we know that Xc_{i_j} is in the span of the first i_j columns of the matrix by definition of h_{id} and thus so are $X^2c_{i_j}, X^3c_{i_j}, X^4c_{i_j}, \ldots$. The pivots of all nonzero elements of $c_{i_j}, Xc_{i_j}, X^2c_{i_j}, X^3c_{i_j}, X^4c_{i_j}, \ldots$ are distinct by construction of X. The pivots for different i_j are distinct because they correspond to generators $t^a e_b$ for different e_b.

This means that $\pi_K(C_{\overline{\tau},\mathcal{H}_{id}})$ contains the K-dimensional space spanned by $c_{i_1}, c_{i_2}, \ldots, c_{i_n}$ together with all images under successive powers of X of those vectors. In other words, $\pi_K(C_{\overline{\tau},\mathcal{H}_{id}})$ is the K-dimensional space $\pi(\Omega_w)$ as desired. □

Proof of Theorem 7.1. Let $Gr(K, N)$, $\mathcal{H}_{id}$, and $\overline{\tau}$ be as in Proposition 7.2.

The map $\pi_K : Fl(N) \to Gr(K, N) = G/P$ restricts to an isomorphism $C_\mu \cong B\mu P/P$ for $\mu \in S_N$ when μ is a Grassmannian permutation so $C_\mu \cong \pi_K(C_\mu)$.

Since $\overline{\tau}$ is a Grassmannian permutation for $Gr(K, N)$ and $\overline{\tau}B$ is contained in $\mathcal{H}_{id}$, it follows that

$$C_{\overline{\tau},\mathcal{H}_{id}} \cong \pi_K\left(C_{\overline{\tau},\mathcal{H}_{id}}\right).$$

This is equal to $\pi(\Omega)$ by Proposition 7.2, so setting $\sigma = \overline{\tau}$ completes the proof. □

Example 7.3. *We describe N and K for Examples 5.2 and 5.3. Both examples show vectors needed for a minimum slice. For notational convenience we pick $N = 6$ in Example 5.2 and $N = 9$ in Example 5.3. The slice corresponding to $N = 6$ in Example 5.2 contains the generator te_2 so $K = 5$, corresponding to the vectors $v_1, v_2, tv_1, v_3, tv_2$. (The last vector is not shown in the example.) By contrast Example 5.3 shows all the generators $t^a e_b$ in the slice and so the column vectors shown in the example are precisely the desired set of vectors, meaning $K = 7$.*

Example 7.4. *We show τ and $\overline{\tau}$ for Example 5.3, with τ on the left. The convention that B is* upper-triangular *means that the standard coset representatives for $B\sigma B/B \in G/B$ have pivots in the* lowest *nonzero entry in each column, with free entries* above *and zeros identically to the right. This reverses the order $\prec$ so the top row of the matrix for a flag corresponds to the largest generator $t^a e_b$ and the bottom row corresponds to the smallest generator.*

$$
\left(\begin{array}{ccccccc|cc}
b_6 & b_{11} & \alpha_{1,4} & \alpha_{2,4} & \alpha_{1,1} & \alpha_{2,1} & 1 & 0 & 0 \\
b_5 & b_{10} & b_3 & b_8 & b_1 & 1 & 0 & 0 & 0 \\
b_4 & b_9 & b_2 & b_7 & 1 & 0 & 0 & 0 & 0 \\
\alpha_{1,4} & \alpha_{2,4} & \alpha_{1,1} & \alpha_{2,1} & 0 & 0 & 0 & 1 & 0 \\
b_3 & b_8 & b_1 & 1 & 0 & 0 & 0 & 0 & 0 \\
b_2 & b_7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\alpha_{1,1} & \alpha_{2,1} & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
b_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right)
\qquad
\left(\begin{array}{ccccccc|cc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \alpha_{2,1} & \alpha_{1,1} & \alpha_{2,4} & \alpha_{1,4} & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \alpha_{2,1} & \alpha_{1,1} & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{array}\right)
$$

References

1. M. Atiyah, I. MacDonald, *Introduction to Commutative Algebra* (Addison-Wesley, Reading, 1969)
2. H. Abe, M. Harada, T. Horiguchi, M. Masuda, The cohomology rings of regular nilpotent Hessenberg varieties in Lie type A, arXiv:1512.09072
3. Darius Bayegan, Megumi Harada, A Giambelli formula for the S^1-equivariant cohomology of type A Peterson varieties. Involve **5**, 115–132 (2012)
4. Sara Billey, Steve Mitchell, Smooth and palindromic Schubert varieties in affine Grassmannians. J. Algebr. Comb. **31**, 169–216 (2010)
5. F. De Mari, C. Procesi, M. Shayman, Hessenberg varieties. Trans. Am. Math. Soc. **332**, 529–534 (1992)
6. Elizabeth Drellich, Monk's rule and Giambelli's formula for Peterson varieties of all Lie types. J. Algebr. Comb. **41**, 539–575 (2015)
7. Lucas Fresse, A unified approach on Springer fibers in the hook, two-row and two-column cases. Trans. Groups **15**(2), 285–331 (2010)

8. Lucas Fresse, Anna Melnikov, On the singularity of the irreducible components of a Springer fiber in $mathfraksl_n$. Sel. Math. (N.S.) **16**(3), 393–418 (2010)
9. W. Fulton, *Tableaux Young, with Applications to Representation Theory and Geometry*, vol. 35, London Mathematical Society Student Texts (Cambridge University Press, Cambridge, 1997)
10. Y.C. Francis, Fung, On the topology of components of some Springer fibers and their relation to Kazhdan–Lusztig theory. Adv. Math. **178**(2), 244–276 (2003)
11. A. Garsia, C. Procesi, On certain graded S_n-modules and the q-Kostka polynomials. Adv. Math. **94**, 82–138 (1992)
12. M. Khovanov, Crossingless matchings and the cohomology of (n, n) Springer varieties. Commun. Contemp. Math. **6**(4), 561–577 (2004)
13. B. Kostant, Flag manifold quantum cohomology, the Toda lattice, and the representation with highest weight ρ. Sel. Math. (N.S.) **2**, 43–91 (1996)
14. B. Kostant, S. Kumar, The nil Hecke ring and cohomology of G/P for a Kac-Moody group G. Adv. Math. **62**, 187–237 (1986)
15. T. Lam, Schubert polynomials for the affine Grassmannian. J. Am. Math. Soc **21**, 259–281 (2008)
16. T. Lam, L. Lapointe, J. Morse, M. Shimozono, Affine insertion and Pieri rules for the affine Grassmannian. Am. J. Math. **128**(6), 1553–1586 (2006)
17. T. Lam, L. Lapointe, J. Morse, A. Schilling, M. Shimozono, M. Zabrocki, *k-Schur functions and affine Schubert calculus*, vol. 33, Fields Institute Monographs (2014)
18. T. Lam, M. Shimozono, Quantum cohomology of G/P and homology of affine Grassmannian. Acta. Math. **204**, 49–90 (2010)
19. L. Lapointe, J. Morse, Quantum cohomology and the k-Schur basis. Trans. Am. Math. Soc. **360**, 2021–2040 (2008)
20. D. Peterson, *Quantum cohomology of G/P* (Lecture notes, Massachusetts Institute of Technology, Cambridge, 1997)
21. M. Precup, The Betti numbers of regular Hessenberg varieties are palindromic. Trans. Groups **23**, 491–499 (2018)
22. K. Rietsch, Totally positive Toeplitz matrices and quantum cohomology of partial flag varieties. J. Am. Math. Soc. **16**, 363–392 (2003)
23. T. Springer, Trigonometric sums, Green functions of finite groups and representations of Weyl groups. Invent. Math. **36**, 173–207 (1976)
24. J. Tymoczko, Linear conditions imposed on flag varieties. Am. J. Math. **128**, 1587–1604 (2006)
25. J. Tymoczko, Paving Hessenberg varieties by affines. Sel. Math. (N.S.) **13**, 353–367 (2007)
26. J. Tymoczko, Permutation representations on Schubert varieties. Am. J. Math. **130**(5), 1171–1194 (2008)
27. X. Zhu, An introduction to affine Grassmannians and the geometric Satake equivalence, arXiv:1603.05593 [math.AG]

A Survey of the Shi Arrangement

Susanna Fishel

Abstract In [58], Shi proved Lusztig's conjecture that the number of two-sided cells for the affine Weyl group of type A_{n-1} is the number of partitions of n. As a byproduct, he introduced the *Shi arrangement* of hyperplanes and found a few of its remarkable properties. The Shi arrangement has since become a central object in algebraic combinatorics. This article is intended to be a fairly gentle introduction to the Shi arrangement, intended for readers with a modest background in combinatorics, algebra, and Euclidean geometry.

In his 1986 paper "The Kazhdan-Lusztig cells in certain affine Weyl groups" [58], Jian-Yi Shi proved Lusztig's conjecture that the number of two-sided cells for the affine Weyl group of type A_{n-1} is the number of partitions of n. As a byproduct, he introduced the *Shi arrangement* of hyperplanes and found a few of its remarkable properties. The Shi arrangement has since become a central object in algebraic combinatorics. This article is intended to be a gentle introduction to the Shi arrangement, intended for readers with a modest background in combinatorics, algebra, and Euclidean geometry. After a review of background material in Sect. 1, we discuss how the Shi arrangement arose in Sect. 2, some of its marvelous enumerative properties in Sect. 3, and some of its surprising connections to algebra in Sect. 4. For some brief comments on recent extensions, see Sect. 5 and for an incomplete list of topics we left out, see Sect. 6.

1 Background

In this section, we will give very brief introductions to some of the ingredients needed to define the Shi arrangement.

S. Fishel (✉)
School of Mathematical and Statistical Sciences, Arizona State University,
P. O. Box 871804, Tempe, AZ 85287-1804, USA
e-mail: sfishel1@asu.edu

H. Barcelo et al. (eds.), *Recent Trends in Algebraic Combinatorics*,
Association for Women in Mathematics Series 16,
https://doi.org/10.1007/978-3-030-05141-9_3

1.1 Root Systems and Coxeter Group Notation

Let V be a finite dimensional real vector space with fixed inner product $\langle | \rangle$. We'll use Δ to denote a *root system*: a finite set of vectors in V which satisfies

(1) $\Delta \cap \mathbb{R}\alpha = \{\alpha, -\alpha\}$ for all $\alpha \in \Delta$ and
(2) $s_\alpha \Delta = \Delta$ for all $\alpha \in \Delta$,

where s_α is the reflection about the hyperplane with normal α. We use Δ^+ to denote a choice of positive roots of Δ, so that $\Delta = \Delta^+ \cup -\Delta^+$, and Π to denote the simple roots, which are a basis for the $\mathbb{R}$-span of Δ. The reflections $S = \{s_\alpha\}_{\alpha\in\Pi}$ generate a finite reflection group W. The *rank* of the system and of W is the dimension of the space spanned by Δ.

Coxeter groups generalize finite reflection groups. Let W be a group with a set of generators $S \subset W$. Let m_{st} be the order of the element st, with $s, t \in S$. If there is no relation between s and t, we set $m_{st} = \infty$. If W has a presentation such that

(1) $m_{ss} = 1$
(2) for $s, t \in S$, $s \neq t$, $1 < m_{st} \leq \infty$,

then W is a *Coxeter group*. We refer to (W, S) as a *Coxeter system*. If $m_{st} \in \{2, 3, 4, 6\}$ when $s \neq t$, then the Coxeter group is called *crystallographic* and, if finite, is a *Weyl* group. It is also a reflection group. The product in any order of all the elements in S is called a *Coxeter element*; all Coxeter elements for a given W are conjugate and their order is the *Coxeter number* of W.

The expression for the reflection s_α, $\alpha \in \Delta$, is

$$s_\alpha(v) = v - 2\frac{\langle v \mid \alpha\rangle}{\langle \alpha \mid \alpha\rangle}\alpha$$

for $v \in V$. For any $k \in \mathbb{R}$, we can define an *affine reflection* $s_{\alpha,k}$ by

$$s_{\alpha,k}(v) = v - 2\frac{\langle v \mid \alpha\rangle - k}{\langle \alpha \mid \alpha\rangle}\alpha.$$

We define the *affine Weyl group* to be the group generated by all affine reflections $s_{\alpha,k}$ for $\alpha \in \Delta$ and $k \in \mathbb{Z}$. It is also a Coxeter group. Its simple reflections are the simple reflections S of the finite Weyl group, plus an extra reflection, s_0, about a translate of a certain other hyperplane in the arrangement. See Fig. 1. Its root system, for us, is the root system for the corresponding finite Weyl group. Our proofs will not be detailed enough to need the full set of affine roots, and we will not define them. Given a root system Δ, we write W_Δ for the corresponding finite group. Please see [41] or [43] for more information.

We put a partial order on any root system. The *root poset* of Δ is the set of positive roots Δ^+, partially ordered by setting $\alpha \leq \beta$ if $\beta - \alpha$ is a nonnegative linear combination of simple roots. If the root system is irreducible, then there is a unique highest root relative to this ordering. We will denote this root by θ. See Fig. 2 for a picture of the root poset for type A_4.

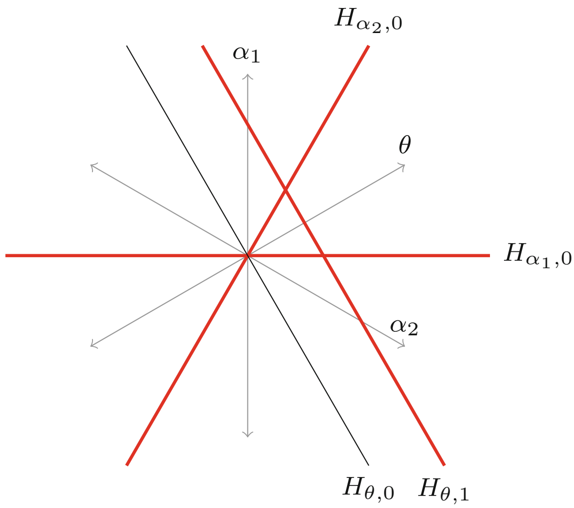


Fig. 1 The roots and reflecting hyperplanes of affine type A_2. The reflection s_1 (resp. s_2) flips the plane over the hyperplane $H_{\alpha_1,0}$ ($H_{\alpha_2,0}$). The reflection s_0 reflects over $H_{\theta,1}$

Let W be a Coxeter group. Every $w \in W$ has an expression as a product of elements of S: $w = s_{i_1} \cdots s_{i_k}$. If k is minimal among all expressions for w, then k is the *length* $\ell(w)$. Any expression for w of length $\ell(w)$ is a *reduced expression*.

We will often refer to the *type* of a group or root system, particularly "type A" and "affine type A," which are the symmetric group or affine symmetric group if we are referring to groups. Please see [41] for more information on the classification of finite reflection groups and Coxeter groups. Humphreys and [17] are good sources for the definitions of irreducible, Bruhat order, and other material omitted here.

1.1.1 Type A

We will be seeing type A often, so we'll be a little more concrete. For A_{n-1}, the vector space V is $\{(a_1, \ldots, a_n) \in \mathbb{R}^n \mid a_1 + \cdots + a_n = 0\}$. Let $\{\varepsilon_1, \ldots, \varepsilon_n\}$ be the standard basis of $\mathbb{R}^n$ and $\langle \mid \rangle$ be the bilinear form for which this is an orthonormal basis. The set

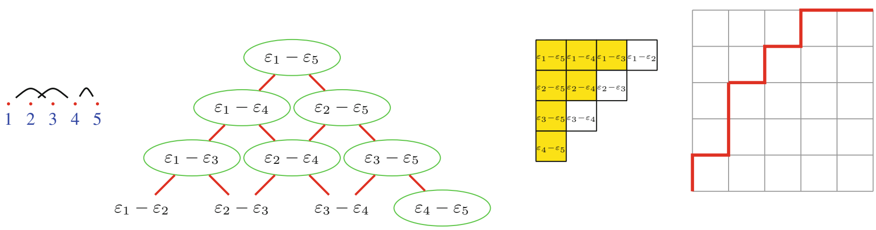


Fig. 2 On the left is the nonnesting set partition $\pi = \{\{1, 3\}, \{2, 4, 5\}\}$. Next to it is the root poset of type A_4 with the filter corresponding to π circled. The third figure is the partition inside the staircase which corresponds to π. On the right is the corresponding Dyck path

of roots is $\Delta = \{\varepsilon_i - \varepsilon_j \mid i \neq j\}$ and a root $\alpha \in \Delta$ is positive, written $\alpha > 0$, if $\alpha \in \Delta^+ = \{\varepsilon_i - \varepsilon_j \mid i < j\}$. Set $\alpha_i = \varepsilon_i - \varepsilon_{i+1}$; the simple roots are $\{\alpha_1, \dots, \alpha_{n-1}\}$. The set Π is a basis of V.

The Coxeter group that Shi studied was the affine symmetric group $\widehat{\mathfrak{S}}_n$, and we review that here. There are several possible descriptions, here we give one due to Lusztig [48]. It is the set of permutations w of $\mathbb{Z}$ such that

(1) $w(i+n) = w(i) + n$ for all $i \in \mathbb{Z}$
(2) $\sum_{i=1}^{n} w(i) = \binom{n+1}{2}$

It's a Coxeter group: for any i, $0 \leq i < n$, s_i corresponds to the permutation

$$t \mapsto \begin{cases} t & \text{if } t \bmod n \neq i \text{ and } t \bmod n \neq (i+1) \bmod \mathrm{n} \\ t-1 & \text{if } t \bmod n = (i+1) \bmod \mathrm{n} \\ t+1 & \text{if } t \bmod n = i \end{cases}$$

The set of reflections S is $\{s_1, \dots, s_{n-1}, s_0\}$ and

$$\widehat{\mathfrak{S}}_n = \langle s_1, \dots, s_{n-1}, s_0 \rangle.$$

The affine symmetric group contains the symmetric group $\mathfrak{S}_n$ as a subgroup. $\mathfrak{S}_n$ is the subgroup generated by the s_i, $0 < i < n$. We identify $\mathfrak{S}_n$ as permutations of $\{1, \dots, n\}$ by identifying s_i with the simple transposition $(i, i+1)$. We act on the right, as did Shi.

1.2 *A Taste of Coxeter Combinatorics, Type A*

The number of parking functions and the Catalan numbers appear in every discussion of the Shi arrangement. We'll define the parking functions when we first see them, in Sect. 3.5, but we'll collect some facts on the Catalan objects here, mostly type A. There are an awful lot of Catalan objects, but only a few of them will appear in this survey.

A *partition* is a finite sequence $\lambda = (\lambda_1, \dots, \lambda_r)$ of positive integers in decreasing order: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$. We identify a partition with its *Young diagram*, the left-justified array of boxes where the ith row from the top has λ_i boxes. A box is called *removable* (respectively *addable*) if we can remove (respectively add) it and still have a diagram of a partition. We use $|\lambda| = \lambda_1 + \dots + \lambda_r$ and $\ell(\lambda) = r$.

1.2.1 Set Partitions

We denote the set $\{1, 2, \dots, n\}$ by $[n]$. The nonempty subsets $B_1, \dots, B_k$ of $[n]$ are a *set partition* of $[n]$ if they are pairwise disjoint and their union is $[n]$. We denote

the set partition $\{B_1, \ldots, B_k\}$ by $B_1|\cdots|B_k$. For example, 13|256|4 is a partition of [6]. The *arc diagram* of a set partition π is defined as follows: place the numbers $1, 2, \ldots, n$ in order on a line and draw an arc between each pair $i < j$ such that

- i and j are in the same block of π, and
- there is no k such that $i < k < j$ and i, k, and j are in the same block.

See Fig. 2.

The partition π has k blocks if and only if it has $n - k$ arcs. This is easy to see if the partition has no arcs. Consider a partition with k blocks and $n - k$ arcs, where i and j are in different blocks. Suppose we add an arc from i to j. We have joined i's and j's blocks, and we now have $k - 1$ blocks and $n - k + 1$ arcs.

A set partition is called *noncrossing* if there does not exist $i < j < k < l$ such that there is an arc from i to k and an arc from j to l. There are C_n noncrossing set partitions, where

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

is the Catalan number (type A). It is called *nonnesting* if there does not exist $i < j < k < l$ such that there is an arc from i to l and an arc from j to k.

1.2.2 Dyck Paths

A *Dyck path* of length n is a lattice path which starts at (0,0), takes only north or east steps of length 1, never goes below the line $y = x$, and ends at (n, n). A north step followed by an east step is called a *valley* of the path.

1.2.3 Root Poset

An *ideal* of a poset P is a subset I of the elements of P such that if $x \in I$ and $y \leq x$ then $y \in I$. A *filter* is like an ideal, except that the condition becomes that $x \in I$ and $x \leq y$ implies $y \in I$. A subset X of the elements of P is an *antichain* if no two elements in X are comparable. An ideal is determined by its maximal elements, which form an antichain, just as a filter is by its minimal elements. These antichains are also called nonnesting partitions; in type A the bijection to the nonnesting partition described above is simply sending the root $\varepsilon_i - \varepsilon_j$ in an antichain to the arc from i to j. There are Catalan number of ideals, filters, and antichains in the root poset defined in Sect. 1.1. We can map a filter F in the root poset for type A_{n-1} to a partition λ whose diagram fits inside the staircase partition $(n-1, n-2, \ldots, 1)$ by the rule that $\varepsilon_i - \varepsilon_j \in F$ if and only if the box in row i and column $n + 1 - j$ is in the diagram of λ. The minimal elements of F correspond to the removable boxes of λ.

There are well-known bijections among all these objects; please see Stanley's book on the subject [71].

1.3 Deformation of Coxeter Arrangements

A *(real) hyperplane arrangement* $\mathcal{H}$ is a set of hyperplanes, possibly affine hyperplanes, in a real vector space. For us, the vector space will be V, the span of some root system Δ, with a fixed inner product $\langle | \rangle$ which is W_Δ invariant. We'll be looking at connected components of a hyperplane arrangement's complement $V \setminus \bigcup_{H \in \mathcal{H}} H$. We will refer to these as the *regions* of the arrangement. The closure $\bar{R}$ of the region R is a convex polyhedron. A *face* of $\mathcal{H}$ is a nonempty set of the form $\bar{R} \cap x$, where x is an intersection of hyperplanes in $\mathcal{H}$. The dimension of a face is the dimension of its affine span. See Stanley [67] for more details. A *wall* H of R is a hyperplane $H \in \mathcal{H}$ such that $\dim(H \cap \bar{R}) = \dim(H)$. The word "bounded" applied to a region has its usual meaning: a region is *bounded* if there is a real number M such that all points in the region are within distance M of the origin. Let $\mathfrak{r}(\mathcal{A})$ and $\mathfrak{b}(\mathcal{A})$ be the number of regions and number of bounded regions, respectively, of the arrangement $\mathcal{A}$.

Let Δ be a root system. The roots (plus the integers) define a system of affine hyperplanes

$$H_{\alpha,k} = \{v \in V \mid \langle v \mid \alpha \rangle = k\}.$$

Note $H_{-\alpha,-k} = H_{\alpha,k}$. In type A, we will sometimes write $x_i - x_j = k$ instead of $H_{\alpha_i + \cdots + \alpha_{j-1},k}$.

The Coxeter arrangement, also called the braid arrangement, is defined

$$\mathcal{C}ox_\Delta = \{H_{\alpha,0} : \alpha \in \Delta^+\}.$$

We give the regions of this arrangement the special name *chambers*. Each chamber corresponds to an element of $W = W(\Delta)$. The *dominant* chamber of V is $\bigcap_{i=1}^{n-1} {H_{\alpha_i,0}}^+$, where ${H_{\alpha_i,k}}^+$ is the half-space $\{v \in V \mid \langle v \mid \alpha \rangle \geq k\}$. the dominant chamber corresponds to the identity of W. It is also referred to as the fundamental chamber in the literature.

The affine Coxeter arrangement is all integer translates of the hyperplanes in $\mathcal{C}ox_\Delta$; that is, it is the whole system of hyperplanes $\{H_{\alpha,k}\}_{\alpha \in \Delta^+, k \in \mathbb{Z}}$. In this arrangement, each region is called an *alcove* and the *fundamental alcove* is A_0, the interior of ${H_{\theta,1}}^- \cap \bigcap_{\alpha \in \Pi} {H_{\alpha,0}}^+$, where θ is the highestroot. A *dominant alcove* is one contained in the dominant chamber.

We also have the m-Catalan arrangement:

$$\mathcal{C}at^m_\Delta = \{H_{\alpha,r} : \alpha \in \Delta, 0 \leq r \leq m\}.$$

Let W be an affine Weyl group and V the vector space spanned by its roots. W acts on V via affine linear transformations, and acts freely and transitively on the set of alcoves. In affine type A_{n-1}, s_i reflects over $H_{\alpha_i,0}$ for $1 \leq i \leq n-1$ and s_0 reflects over $H_{\theta,1}$, where θ is the highest root. We identify each alcove A with the unique $w \in W$ such that $A = A_0 w$. For example, if w is the element of affine A_2

$$\mathcal{C}at^m_\Delta = \{H_{\alpha,r} : \alpha \in \Delta, 0 \le r \le m\}.$$

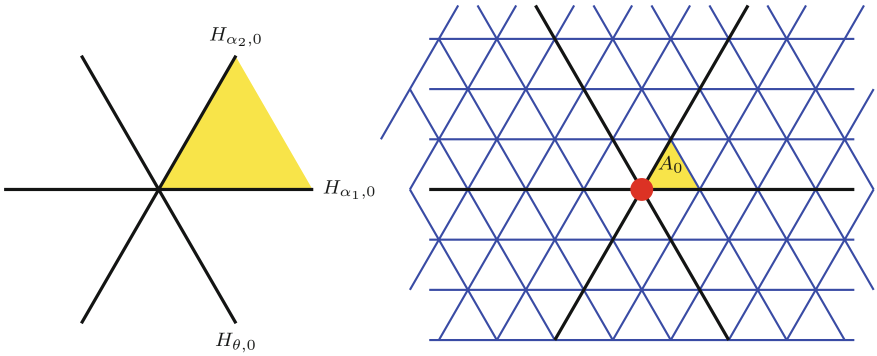


Fig. 3 The Coxeter arrangement for A_2 is on the left, with shaded fundamental chamber. On the right is all its translates, with fundamental alcove shaded

whose reduce decomposition is s_0s_1, then A_0w is the image of A_0 after reflecting first across $H_{\theta,1}$ and then across $H_{\alpha_1,0}$. See Fig. 6.

We can be even more specific for type A_{n-1}. The action on V is given by

$$s_i(a_1,\ldots,a_i,a_{i+1},\ldots,a_n) = (a_1,\ldots,a_{i+1},a_i,\ldots,a_n) \quad \text{for } i \ne 0, \text{ and}$$
$$s_0(a_1,\ldots,a_n) = (a_n+1, a_2,\ldots,a_{n-1},a_1-1).$$

Note $\widehat{\mathfrak{S}}_n$ preserves $\langle \mid \rangle$, but $\widehat{\mathfrak{S}}_n$ does not.

An alcove A can be described by the hyperplanes it is between. For example, in Fig. 6, the alcove labeled by 20 is between $H_{\alpha_1,1}$ and $H_{\alpha_1,2}$, between $H_{\alpha_2,0}$ and $H_{\alpha_2,1}$ and between $H_{\theta,1}$ and $H_{\theta,2}$. Given a positive root α, there is a unique integer $k = k_\alpha(A)$ such that $k < \langle\alpha \mid x\rangle < k+1$ for all $x \in A$. Let $K(A) = \{k_\alpha(A)\}_{\alpha\in\Delta^+}$ denote the set of coordinates for A, indexed by the positive roots. Returning to Fig. 6, the alcove labeled by 20 has coordinates $k_{\alpha_1} = k_\theta = 1$ and $k_{\alpha_2} = 0$.

Shi characterized the sets of integers which can arise as $K(A)$ for some alcove A; for type A in [58, Chapter 6] and for general affine Weyl groups in [59]. The situation in general is rather messy, but if we assume our root system is an irreducible crystallographic one, then Shi found (see also [11]) that a collection of integers indexed by the positive roots Δ^+ corresponds to an alcove if and only if

$$k_\alpha + k_\beta \le k_{\alpha+\beta} \le k_\alpha + k_\beta + 1 \tag{1.1}$$

for all $\alpha, \beta \in \Delta^+$ such that $\alpha+\beta \in \Delta^+$. We call the set $K(A)$ the coordinates of A (Fig. 3).

2 Origin

2.1 Kazhdan–Lusztig Cells

We provide a bare-bones introduction to Kazhdan–Lusztig theory. For more information, see [17, 44, 58]. If you are willing to believe that Kazhdan and Lusztig defined an equivalence relation on the elements of a Coxeter group, then skip this section. We include this collection of definitions for completeness. We'll need the Hecke algebra $\mathfrak{H}$ and the Kazhdan–Lusztig polynomials in order to define the W-graph and then the cells. We will prove none of our claims.

Let W be a Coxeter group and let S be the corresponding set of simple reflections. We first follow [44], who follow [19], for the definition of the Hecke algebra. Let $\mathcal{A}$ be the ring of Laurent polynomials in the indeterminate $q^{1/2}$ with integral coefficients. The Hecke algebra $\mathfrak{H} = \mathfrak{H}(W, S)$ is a free module over $\mathcal{A}$ with basis T_w, one for each $w \in W$. The multiplication is defined by the rules

(1) $T_w T_{w'} = T_{ww'}$ if $\ell(ww') = \ell(w) + \ell(w')$
(2) $(T_s + 1)(T_s - q) = 0$ if $s \in S$;

here $\ell(w)$ is the length of w.

Now for the polynomials. The involution on $\mathcal{A}$ $a \mapsto \bar{a}$ defined by $\overline{q^{1/2}} = q^{-1/2}$ extends to an involution of the ring $\mathfrak{H}$:

$$\overline{\sum a_w T_w} = \sum \overline{a_w} T_{w^{-1}}^{-1}.$$

Kazhdan's and Lusztig's theorem, Theorem 2.1 in this survey, asserts the existence of elements $C_w \in \mathfrak{H}$, one for each $w \in W$, and simultaneously defines the Kazhdan–Lusztig polynomials $P_{y,w}$, where $y, x \in W$. The order is the Bruhat order on W.

Theorem 2.1 ([44]). For any $w \in W$, there is a unique element $C_w \in \mathfrak{H}$ such that

- $\overline{C}_w = C_w$
- $C_w = \sum_{y \leq w} (-1)^{\ell(y)+\ell(w)} q^{\ell(w)/2-\ell(y)} \overline{P_{y,w}} T_y$

where $P_{y,x} \in \mathcal{A}$ is a polynomial in q of degree at most $\frac{1}{2}(\ell(w) - \ell(y) - 1)$ for $y < w$, and $P_{w,w} = 1$.

Kazhdan and Lusztig used the polynomials to define a graph and from there the cells. Now we follow the exposition given in [17], simplified just a bit because we will not prove anything. For $u, w \in W$, define $\mu(u, w)$ to be the coefficient of $q^{\frac{1}{2}(\ell(w)-\ell(u)-1)}$ in $P_{u,w}(q)$ if $u < w$ and $\frac{1}{2}(\ell(w) - \ell(y) - 1)$ is an integer; otherwise, $\mu(u, w)$ is 0. Let $\mathcal{L}(w)$ be the set of left descents of w: $\mathcal{L}(w) = \{s \in S | sw < w\}$. The directed, labeled graph $\tilde{\Gamma}^L_{(W,S)}$ is the graph with vertices $x \in W$ and edges $x \xrightarrow{(s,\mu)} y \in E$. There are two types of edges in E:

(1) $x, y \in W$, $x \neq y$, either $\mu(x, y) \neq 0$ or $\mu(y, x) \neq 0$, and $s \in \mathcal{L}(x) \setminus \mathcal{L}(y)$. Let μ be either $\mu(x, y)$ or $\mu(y, x)$, whichever is not 0.
(2) Loops at x: labeled by $s \in S$ and

$$\mu = \begin{cases} 1 & \text{if } s \notin \mathcal{L}(x), \\ -1 & \text{if } s \in \mathcal{L}(x). \end{cases}$$

The graph $\tilde{\Gamma}^R_{(W,S)}$ has an analogous definition, using right descents of w: $\mathcal{R}(w) = \{s \in S | ws < w\}$. The graph $\tilde{\Gamma}^{LR}_{(W,S)}$ is the superposition of the $\tilde{\Gamma}^L_{(W,S)}$ and $\tilde{\Gamma}^R_{(W,S)}$. We describe the cells in graph theoretic terms. A directed graph is *strongly connected* if there is a directed path between all pairs of vertices. A *strongly connected component* of a directed graph is a maximal strongly connected subgraph. Finally, the *left cells* are the strongly connected components of $\tilde{\Gamma}^L_{(W,S)}$, the *right cells* the strongly connected components of $\tilde{\Gamma}^R_{(W,S)}$, and the *two-sided cells* the strongly connected components of $\tilde{\Gamma}^{LR}_{(W,S)}$.

The list of areas in math where cells appear is mind-boggling. Please see the short survey by Gunnells [32], for example, for references, as well as for insight into the geometry of the cells. The book by Björner and Brenti [17] explains much of the combinatorial connection.

2.2 Shi Regions and Kahzdan–Lusztig Cells

Shi was studying cells in [58]. He concentrated on the affine Weyl groups of type A, because of the following conjectures of Lusztig. In [48], Lusztig defined a map σ from $\widehat{\mathfrak{S}}_n$ to partitions of n. He conjectured that for any partition λ of n, $\sigma^{-1}(\lambda)$, a set of affine permutations, is in fact a two-sided cell. Lusztig also conjectured a formula for the number of left (or right) cells which make up the two-sided cell $\sigma^{-1}(\lambda)$. The description of the map is simple enough and we define here. Let $w \in \widehat{\mathfrak{S}}_n$ and define $d_k = d_k(w)$ to be the maximum size of a subset of $\mathbb{Z}$ whose elements are noncongruent to each other modulo n and which is a disjoint union of k subsets each of which has its natural order reversed by w. The partition λ is given by $(d_1, d_2 - d_1, \ldots, d_n - d_{n-1})$.

Example 2.2. Let $n = 3$. The permutation s_0 is

$$\begin{pmatrix} \cdots & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & \cdots \\ \cdots & -2 & -3 & -1 & 1 & 0 & 2 & 4 & 3 & 5 & 3 & \cdots \end{pmatrix}.$$

The set $\{3, 4\}$ has its order reversed by s_0 and there is no larger set, so $d_1 = 2$. The sets $\{3, 4\}$ and $\{2\}$ show that $d_2 = 3$. Therefore $\sigma(s_0) = (2, 1)$. In our notation, $s_0 s_2$ is

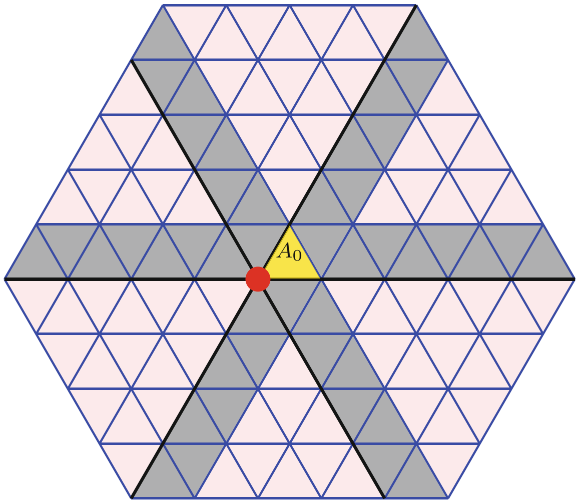


Fig. 4 The cells for affine A_2. The yellow region is the two-sided cell $\sigma^{-1}((1, 1, 1))$, which is also a single left-cell. The six pink regions are left-cells, whose union is the two-sided cell $\sigma^{-1}((3))$. The three gray regions are also left cells, and their union is the two-sided cell $\sigma^{-1}((2, 1))$. See [58, Page 98]

$$\begin{pmatrix} \cdots & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & \cdots \\ \cdots & -2 & -4 & 0 & 1 & -1 & 3 & 4 & 2 & 6 & 7 & \cdots \end{pmatrix}.$$

and $\sigma(s_0 s_2) = (2, 1)$ also. The identity maps to $(1, 1, 1)$ under σ and $s_1 s_2 s_1$, for example, maps to (3).

Shi proved both of Lusztig's conjectures, and more. Shi used the identification of $\widehat{\mathfrak{S}}_n$ with alcoves to describe the cells of affine type A. He showed that the two-sided cells correspond to connected sets of alcoves, one set of alcoves for each partition λ of n. A two-sided cell is a disjoint union of left-cells. Inside the two-sided cell corresponding to the partition λ, there is one left-cell for each tabloid of shape λ. See Fig. 4.

What came to be known as the Shi arrangement was not initially defined in terms of hyperplanes. Shi began by defining *rank n sign types* as triangular arrays $X = (x_{ij})_{1 \le i < j \le n}$ with entries from $\{+, -, \bigcirc\}$. The *admissible sign types* correspond to the regions of his arrangement. He defined them as the sign types which satisfy the following condition: for all $1 \le i < t < j \le n$, the triple

$$\begin{smallmatrix} & x_{ij} & \\ x_{it} & & x_{tj} \end{smallmatrix}$$

is a member of the set G_A of admissible sign types of rank 3 (A_2). G_A is the set

$$\begin{aligned} \{ & \begin{smallmatrix} & + & \\ + & & + \end{smallmatrix}, \begin{smallmatrix} & + & \\ + & & \bigcirc \end{smallmatrix}, \begin{smallmatrix} & + & \\ \bigcirc & & + \end{smallmatrix}, \begin{smallmatrix} & + & \\ + & & - \end{smallmatrix}, \begin{smallmatrix} & + & \\ - & & + \end{smallmatrix}, \begin{smallmatrix} & + & \\ \bigcirc & & \bigcirc \end{smallmatrix}, \begin{smallmatrix} & \bigcirc & \\ \bigcirc & & \bigcirc \end{smallmatrix}, \begin{smallmatrix} & \bigcirc & \\ + & & - \end{smallmatrix}, \\ & \begin{smallmatrix} & \bigcirc & \\ - & & + \end{smallmatrix}, \begin{smallmatrix} & \bigcirc & \\ \bigcirc & & - \end{smallmatrix}, \begin{smallmatrix} & \bigcirc & \\ - & & \bigcirc \end{smallmatrix}, \begin{smallmatrix} & - & \\ - & & - \end{smallmatrix}, \begin{smallmatrix} & - & \\ + & & - \end{smallmatrix}, \begin{smallmatrix} & - & \\ - & & + \end{smallmatrix}, \begin{smallmatrix} & - & \\ - & & \bigcirc \end{smallmatrix}, \begin{smallmatrix} & - & \\ \bigcirc & & - \end{smallmatrix} \ \} \end{aligned} \tag{2.1}$$

Two comments on G_A. If we order the symbols $\{\bigcirc, +, -\}$ as $- < 0 < +$, then G_A can be seen as the rank 3 sign types where either $x_{12} \leq x_{13} \leq x_{23}$ or $x_{23} \leq x_{13} \leq x_{12}$, together with $x_{13} = +, x_{12} = x_{23} = 0$. The set G_A has cardinality 16, which is $(n+1)^{n-1}$ for $n = 3$.

Shi connected the admissible sign types to geometry using (1.1) and the map ζ. If K is the set of coordinates for an alcove A, then define the sign type $X = \zeta(A)$ by

$$x_{ij} = \begin{cases} + & \text{if } k_{ij} > 0 \\ \bigcirc & \text{if } k_{ij} = 0 \\ - & \text{if } k_{ij} < 0. \end{cases}$$

He then calculated the hyperplanes so that the regions defined by them were made up of alcoves with the same image under the map ζ. We use admissible sign type, region in the Shi arrangement, and Shi region interchangeably.

Shi showed in [58] that the left-cells for affine type A are themselves disjoint unions of admissible sign types. Admissible sign types were not used directly in the proofs of the Lusztig conjectures in Shi's monograph, but describe the structure of the cells. They have taken on a life of their own.

Later, in [60], Shi extended the definition of admissible sign types, thereby generalizing the Shi arrangement. This is the definition we give below.

We give the definition for any irreducible, crystallographic root system Δ. When the root system is type A_{n-1}, we will sometimes write $\mathcal{S}hi_n$ instead of $\mathcal{S}hi_\Delta$.

Definition 2.3. The Shi arrangement $\mathcal{S}hi_\Delta$ is the collection of hyperplanes

$$\{H_{\alpha,k} \mid \alpha \in \Delta^+, 0 \leq k \leq 1\}.$$

Example 2.4. All alcoves in the region labeled by

$$\begin{matrix} & \bigcirc & \\ + & & - \end{matrix}$$

in Fig. 5 have positive coordinate $k_{12} = k_{\alpha_1}$, have negative coordinate k_{α_2}, and have the coordinate k_θ equal to zero. Likewise, all three coordinates of all alcoves in the region labeled by

$$\begin{matrix} & + & \\ + & & + \end{matrix}$$

are positive. See Sect. 1.3 for the definition of coordinates of an alcove.

We mention here that in the case where the Coxeter graph of the system contains an edge with a label greater than 3, it is not true that all the left-cells of the affine Weyl group are unions of admissible sign types. It may be conjectured that it holds for any affine Weyl group of simply-laced type. The cells in affine D_4 have been explicitly described by Shi in [63], so the conjecture may not be difficult to verify. It

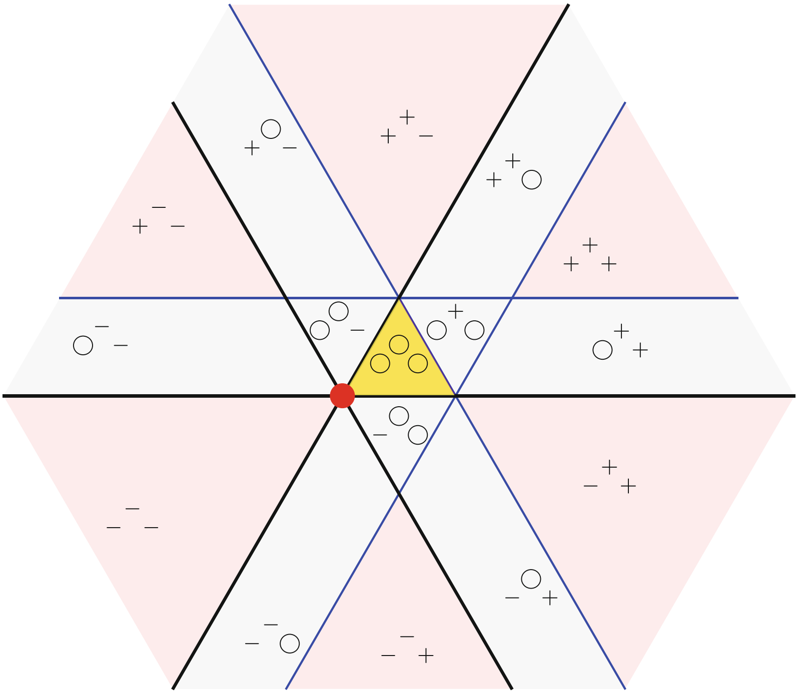


Fig. 5 The Shi arrangement for type A_2. See [58, Page 102]. Each region has been labeled with its sign type. See Example 2.4

is known that any left-cell in the lowest or highest two-sided cell of any irreducible affine Weyl group forms a single admissible sign type; see Shi [61, 62]. We thank Jian-Yi Shi for this information.

By a *dominant region* of the Shi arrangement, we mean a connected component of the hyperplane arrangement complement $V \setminus \bigcup_{H \in \mathcal{S}hi_{\Delta}} H$ that is contained in the dominant chamber. Both the formula for the number of regions in the whole arrangement and for the number of dominant regions are intriguing and will be discussed in the next section.

3 Enumeration

The Shi regions have been counted multiple times. We discuss four different approaches to enumerating them.

3.1 The Number of Shi Regions, Part 1

Shi concentrated on the admissible sign types in Chapter 7 of his book [58], where he introduced the arrangement for type A.

He enumerates them for type A by considering the alcove closest to the origin in each region. We'll call this the *minimal alcove* of the region and denote it A_R if the region is R. Shi called such an alcove the shortest alcove [58, Section 7.3]. He characterized A_R using left descents. Left descents are key to the definition of cells, so it is not surprising that they appear in the description of minimal alcoves. Basically, an alcove is minimal if any reflection which brings it closer to the origin flips it out of the region.

Example 3.1. See Fig. 6. Let $w = s_1s_2s_1s_0$. This w has two left descents, s_1 and s_2, and both $s_1w = s_2s_1s_0$ and $s_2w = s_1s_2s_0$ are in different regions the minimal permutation w. On the other hand, $w = s_1s_2s_0$ is not minimal and indeed $\ell(s_1w) < \ell(w)$ and $s_1w = s_2s_0$ is in the same region as w.

Every alcove corresponds to an affine permutation. We'll call the affine permutations whose alcoves are minimal *minimal permutations*. Shi showed that the collection of alcoves corresponding to the inverses of minimal permutations is exactly a scaled version of the fundamental alcove. See Fig. 6. Thus to calculate the number of regions in his newfound arrangement, he calculated the number of alcoves in this scaled fundamental alcove. He calculated something a bit more general: if the fundamental alcove is expanded by the positive integer m, then it is made up of $m^{\dim(V)}$ alcoves. The alcoves corresponding to the inverses of minimal permutation land in the fundamental alcove scaled by $n + 1$, which showed that there are $(n + 1)^{n-1}$ regions in the Shi arrangement of type A. The expression $(n + 1)^{n-1}$ pops up frequently in combinatorics and algebra; see [35] for their connection to q, t-Catalan numbers, for example.

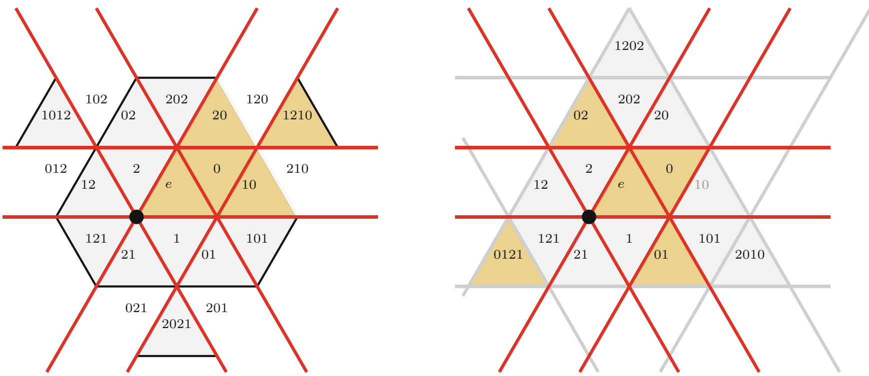


Fig. 6 On the left, the minimal alcove of each Shi region is shaded. The dominant ones are shaded yellow, the rest gray. We've labeled the alcoves by the corresponding affine permutation, where we use i for s_i for space reasons. The alcoves corresponding to the inverses are drawn on the right

Shi's enumeration of the regions is perhaps more complicated than the others described here. However, his discovery that the inverses of the minimal permutations correspond to a simplex is worth the price of admission. The minimal alcoves have been useful in other enumeration; see [12, 28], for example. For another example, Hohlweg, Nadeau, and Williams, in [39], generalize the Shi arrangement to any Coxeter group (and beyond!) and conjecture that the inverses of the analogues of minimal permutations form a convex body. See also Sommers [66] where the simplex was generalized to what is now called the Sommers region. Thomas and Williams [75] show that the set of alcoves in this region, and by extension the Shi regions, exhibit the cyclic sieving phenomenon.

In [60], Shi generalized sign types to other affine Weyl groups. He defined sets analogous to (2.1) for other types. The hyperplane arrangements were still given by Definition 2.3, but now he considered root systems other than type A. He used the map ζ on alcoves and described the sign types which arose and again characterized the element in each region with the minimal number of hyperplanes separating it from the origin. As above, we identify elements $w \in W$ with $A_0 w$ and refer to w as minimal if its alcove is minimal. The fact that

$$\bigcup_{w \text{ minimal}} A_0 w^{-1}$$

is a simplex is not just a type A phenomenon. Shi proved it for other affine Weyl groups and used it to prove that there are $(h+1)^n$ regions, where h is the Coxeter number of the system.

Shi counted the number of regions in the dominant chamber for affine Weyl groups in [64]. He calls the admissible sign types corresponding to regions in the dominant chamber $\oplus$-sign types. For types A, B, C, and D, he finds a bijection from $\oplus$-sign types to filters in the root poset for Δ^+. Here is a technical detail: Shi finds the bijection to the positive coroots $(\Delta^+)^\vee$, which we won't define, then mentions that it has the same type as Δ^+ except when Δ^+ has type B and C. He deals with types B and C separately. We will continue using Δ^+. He further maps the filters to subdiagrams of certain Young diagrams; see Sect. 1.2.3. For example, for type A, the subdiagrams are those of partitions whose diagrams fit inside the staircase shape. In the exceptional cases, he enumerates increasing subsets directly. He shows the $\oplus$-sign types for affice Weyl groups are enumerated by the Catalan numbers, although Shi does not mention them.

A few more words are in order on this important bijection to filters in the root poset. The key proposition, from Section 1.2 of [64], follows (using roots instead of coroots).

Proposition 3.2. Assume that $X = (X_\alpha)_{\alpha \in \Delta^+}$ is a Δ^+-tuple with $X_\alpha \in \{+, \bigcirc\}$. Then X is an $\oplus$-sign type if and only if the following condition on X holds: if $\alpha, \beta \in \Delta^+$ satisfy $\beta > \alpha$ and $X_\alpha = +$, then $X_\beta = +$.

We'll use type A as an example. In the set G_A in (2.1), there are five triples which contain only $\bigcirc$ and $+$, but there are eight which are possible. The condition

in Proposition 3.2 rules the other three out, proving sufficiency. For necessity, Shi uses induction to reduce to the rank two case and shows that the five $\oplus$-sign types in G satisfy the condition.

3.2 Interlude

We'll need these standard definitions for Sects. 3.3 and 3.4. See [70] for the definitions of the rank function ρ and Möbius function μ of a poset.

Definition 3.3. [70] Let P be a finite graded poset with $\hat{0}$. Let ρ be its rank function and n the rank of P. Define the characteristic polynomial $\chi_P(x)$ of P by

$$\chi_P(x) = \sum_{t \in P} \mu(\hat{0}, t) x^{n-\rho(t)}.$$

Definition 3.4. [70] Let $\mathcal{A}$ be a hyperplane arrangement in a vector space V and let $L(\mathcal{A})$ be the set of all nonempty intersections of hyperplanes in $\mathcal{A}$. Include V itself, by considering it as the intersection over the empty set. Order $L(\mathcal{A})$ by reverse inclusion, so that $\hat{0}$ is V.

See Fig. 7 for an example of the poset of intersections.

If the intersection of all the hyperplanes in $\mathcal{A}$ is nonempty, then $L(\mathcal{A})$ is a lattice. The intersection of all hyperplanes in $\mathcal{S}hi_\Delta$ is empty and $L(\mathcal{S}hi_\Delta)$ will only be a meet semi-lattice. It is finite and graded by $\rho(t) = n - \dim(t)$, where $n = \dim(V)$. The characteristic polynomial of an arrangement is

$$\chi_{\mathcal{A}}(x) = \sum_{t \in L(\mathcal{A})} \mu(\hat{0}, t) x^{n-\rho(t)}$$

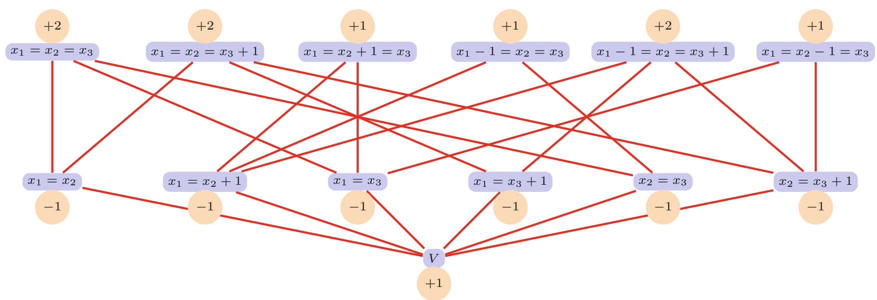


Fig. 7 The poset of intersections for type A_2. For example, the node labeled $x_1 = x_2 = x_3 + 1$ represents $H_{\alpha_1,0} \cap H_{\alpha_2,0} \cap H_{\alpha_3,1}$. See [37, Page 36]. The numbers in the circle above or below each element is the Möbius function of the interval from the element to $V = \hat{0}$. The Poincaré polynomial is $P_{\Delta_{A_2}}(x) = (1 + 3x)^2$

and the Poincaré polynomial is

$$P_{\mathcal{A}}(x) = \sum_{t \in L(\mathcal{A})} \mu(\hat{0}, t)(-x)^{\rho(t)}.$$

The Poincaré polynomial is a rescaled version of the characteristic polynomial.

The characteristic polynomial is invaluable for studying hyperplane arrangements, thanks to a theorem of Zaslavsky [76, 77]. See also Stanley's notes on hyperplanes [69]. In Sect. 1.3, we defined $\mathfrak{r}(\mathcal{A})$ and $\mathfrak{b}(\mathcal{A})$ be the number of regions and number of bounded regions of the arrangement $\mathcal{A}$.

Theorem 3.5 ([77])**.** Let $\mathcal{A}$ be an arrangement in an n-dimensional real vector space. Then

$$\mathfrak{r}(\mathcal{A}) = (-1)^n \chi_{\mathcal{A}}(-1) = P_{\mathcal{A}}(1)$$
$$\mathfrak{b}(\mathcal{A}) = (-1)^{\operatorname{rank} \mathcal{A}} \chi_{\mathcal{A}}(1) = P_{\mathcal{A}}(-1).$$

3.3 The Number of Shi Regions, Part 2

Headley [37, 38] calculated the Poincaré polynomial $P_{\mathcal{S}hi_\Delta}$ of the Shi arrangement for an irreducible root system. He found a recursion for its coefficients, which we will now present.

Let $\overline{\mathcal{S}hi_\Delta}$ be the subarrangement of $\mathcal{S}hi_\Delta$ consisting of all hyperplanes which contain the origin. For $Y \in L(\overline{\mathcal{S}hi_\Delta})$, let W_Y be the group generated by the reflection through all the hyperplanes containing Y. For a polynomial $p(t)$, let $[t^k]p(t)$ be the coefficient of t^k in $p(t)$.

Lemma 3.6 ([38])**.** For $Y \in L(\overline{\mathcal{S}hi_\Delta})$, let $W_{Y,1} \times \cdots W_{Y,m}$ be the decomposition of W_Y into irreducible Coxeter groups. Let $S_i = \mathcal{S}hi_{W_{Y,i}}$ be the Shi arrangement associated to the Coxeter group $W_{Y,i}$. Then

$$[t^k]P_{\mathcal{S}hi_\Delta}(t) = [t^k] \sum_{Y \in L(\overline{\mathcal{S}hi_\Delta}):\operatorname{rank}(Y)=k} P_{S_1}(t) \cdots P_{S_m}(t).$$

Headley uses induction on the number of generators to determine every coefficient except the leading one. For this, he relies on Shi's enumeration and the relationship between the Poincaré polynomial and the number of regions. His analysis is done case by case for each Coxeter type. His theorem is

Theorem 3.7. Let Δ be an irreducible crystallographic root system, with Coxeter number h and rank n. Then

$$P_{\mathcal{S}hi_\Delta}(t) = (1 + ht)^n.$$

In type A, the argument for calculating the coefficients is simple enough to repeat: he matches an element Y of the intersection poset $L(\hat{\mathcal{S}hi}_\Delta)$ with the set partition $B = (B_1, \ldots, B_m)$ of $[n+1]$ by

$$Y = \cap\{x_i - x_j = 0 : i, j \text{ are in the same block of } B\}.$$

In this case, W_Y is isomorphic to $A_{|B_1|-1} \times \cdots \times A_{|B_m|-1}$ and $\operatorname{rank}(Y) = n+1-m$. Therefore, by induction and Lemma 3.6,

$$[t^k]P_{\mathcal{S}hi_\Delta}(t) = \sum_{\text{Partitions of } [n+1] \text{ into } n+1-k \text{ blocks}} |B_1|^{|B_1|-1} \cdots |B_{n+1-k}|^{|B_{n+1-k}|-1}. \tag{3.1}$$

In his thesis [37], he used Lagrange inversion to calculate

$$[t^k]P_{\mathcal{S}hi_\Delta} = (n+1)^k \binom{n}{n-k}.$$

Later, in [38], he recognized the sum in (3.1) to be the number of labeled forests on $n+1$ vertices of $n+1-k$ trees and used [51]. In both his thesis and later paper, he showed that the coefficient of t^k in $P_{\mathcal{S}hi_\Delta}(t)$ and $(1+(n+1))^n$ are the same for $1 \leq k \leq n-1$. Then since the degree of $P_{\mathcal{S}hi_\Delta}(t)$ is n and since $P_{\mathcal{S}hi_\Delta}(1) = (n+2)^n$ by Shi's result, he showed

$$P_{\mathcal{S}hi_\Delta}(t) = (1+(n+1)t)^n = (1+ht)^n \tag{3.2}$$

in A_n.

3.4 The Number of Shi Regions, Part 3

Crapo and Rota [23, Chapter 16] described the *critical problem*: let S be a set of points in an n-dimensional vector space V_n over the field $\mathbb{F}_q$ with q elements. The set S must not contain the origin. Find the minimum number c of projective hyperplanes $H_1, \ldots, H_c$ with the property that the intersection $H_1 \cap \ldots H_c \cap S$ is null. They were able to solve the problem using the poset of intersections and characteristic polynomial.

Athanasiadis [8] turned Crapo and Rota's theorem around and used it to calculate the characteristic polynomial of subspace arrangements. Blass and Sagan [18] had previously used a similar idea, but not for all subspaces and not for the Shi arrangement. We present first the the Crapo and Rota theorem, then describe how Athanasiadis used it to get his hands on the characteristic polynomial for the Shi arrangements for irreducible crystallographic root systems.

Theorem 3.8 ([23]). The number of linearly ordered sequences $(L_1, L_2, \ldots, L_k)$ of k linear functionals in V_n which distinguish the set S is given by $p(q^k)$ where $p(v)$ is the characteristic polynomial of the geometric lattice spanned by the set S.

Athanasiadis needed to count the number of n-tuples $(x_1, \ldots, x_n) \in \mathbb{F}_q^n$ which satisfy $x_i \neq x_j$ and $x_i \neq x_j + 1$ for $i < j$. The argument is simple (and lovely) enough in type A_{n-1} for the full Shi arrangement that we reproduce it here. See also [70].

We first solve a related problem. Find the number of ways there are to place n labeled balls in q unlabeled boxes, where

(1) the boxes are in a circle,
(2) there is never more than one ball in a box, and
(3) if $i < j$, then ball i is not placed in the box immediately following, in the clockwise direction, the box holding ball j.

There will be $q - n$ empty boxes, so first place them in a circle. There is one way to do that. There are now $q - n$ spaces between the empty boxes, where the boxes holding the balls will go. By cyclic symmetry, there is one way to place the box holding the 1-ball. Then there are $(q - n)^{n-1}$ ways to place the rest of the boxes holding balls in the empty spaces. It is enough to pick the space between empty boxes: to avoid violating condition (3), the boxes between a consecutive pair of empty boxes must placed in increasing order of the labels on the balls inside. That is, our final answer to the related problem is $(q - n)^{n-1}$.

Now back to counting n-tuples. We are essentially done, if we think of each n-tuple representing a distribution of n labeled balls into a circle of q labeled boxes, where the distribution satisfies conditions (2) and (3). We place the ball labeled i in box x_i. We need only label the boxes, and there are q ways to do this. Thus there are $q(q - n)^{n-1}$ n-tuples which satisfy $x_i \neq x_j$ and $x_i \neq x_j + 1$ for $i < j$ and we have that the characteristic polynomial is $\chi_{L(\mathcal{S}hi_n)} = q(q - n)^{n-1}$.

Crapo and Rota's finite field method has since been used to calculate other characteristic polynomials. See Armstrong [5], Armstrong and Rhoades [7], and Ardila [3], for example. See Athanasiadis [13] for reciprocity results for the characteristic polynomial for the Shi arrangement.

3.5 The Number of Shi Regions, Part 4

Pak and Stanley, in [67], give a bijection from Shi regions (type A) to parking functions, which refines (3.2). It is proved to be a bijection in [69, Lecture 6]. A *parking function of length n* is a tuple of nonnegative integers $(p_1, \ldots, p_n)$ such that when rearranged in nondecreasing order and relabeled as $b_1 \leq b_2 \leq \cdots \leq b_n$, then $b_i \leq i - 1$. Parking functions generalize inversion vectors of permutations. Pak and Stanley recursively defined the label $\lambda(R)$ of a region R. We use the description given in [69, Lecture 6].

Let R_0 be the fundamental alcove A_0. Set $\lambda(R_0) = (0, \cdots, 0)$. Suppose we have labeled the region R and its label $\lambda(R)$ is $(a_1, \ldots, a_n)$.

- If the regions R and R' are separated by the single hyperplane H with the equation $x_i - x_j = 0$, $i < j$, and if R and R_0 lie on the same side of H, then $\lambda(R') = (a_1, \ldots, a_{i-1}, a_i + 1, a_{i+1} \ldots, a_{j-1} a_j, a_{j+1}, \ldots, a_n)$.
- If the regions R and R' are separated by the single hyperplane H with the equation $x_i - x_j = 1$, $i < j$, and if R and R_0 lie on the same side of H, then $\lambda(R') = (a_1, \ldots a_{i-1}, a_i, a_{i+1} \ldots, a_{j-1}, a_j + 1, a_{j+1}, \ldots, a_n)$.

The bijection generalizes the well-known bijection from permutations to inversion vectors [70, Chapter 1]. Although the map λ is simply stated, the proof that it is a bijection is not simple. To show that the labeling is a bijection, Stanley encodes each region as a permutation and antichain pair. He builds the inverse map step-by-step from the parking functions to the pairs. The summary by Armstrong [5, Theorem 3] of the proof that the Pak–Stanley map is bijective is particularly good. Recall that the filters/antichains/ideals in the root poset for type A correspond to partitions in a staircase, and define the non-inversions of a permutation w to be the pairs (i, j) such that $i < j$ and $w(i) < w(j)$. Then the proof can be summarized as showing that the Shi regions are in bijection with pairs $(w, \mathcal{I})$ where $w \in \mathfrak{S}_n$ and $\mathcal{I}$ is an ideal in the root poset Δ^+ such that the minimal elements of $\mathcal{I}$, which are labels in the valleys of the Dyck path corresponding to $\mathcal{I}$, are non-inversions of w (Fig. 8).

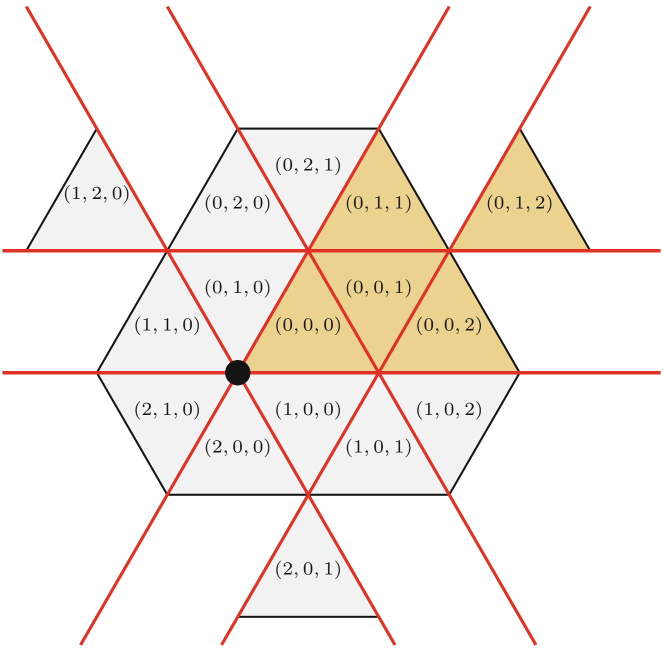


Fig. 8 The Shi regions labeled by parking functions, using the Pak–Stanley labeling. A label is nondecreasing if and only if the region is in the dominant region

The Pak and Stanley bijection from regions to parking functions ($m = 1$) case can be composed with a bijection from trees to parking functions. The number of regions R for which i hyperplanes separate R from the region R_0 is equal to the number of trees on the vertices $0, \ldots, n$ with $\binom{n}{2} - i$ inversions. The pair (i, j), where $1 \leq i < j$, is an inversion for T if the vertex j lies on the unique path in T from 0 to i. See [67, Theorem 5.1].

We mention a few papers which build on the Pak–Stanley bijection. Duertes and Guedes de Oliveira, for example, further analyze this bijection in [25]. Rincón [56] extends the Pak–Stanley labeling to the poset of faces of the Shi arrangement. See also Sect. 5.

3.6 More

Believe it or not, there are still other wonderful proofs concerning the number of regions.

For example, Athanasiadis and Linusson [14] defined a bijection, different from Pak and Stanley's, from parking functions to the Shi regions (type A). Theirs gives a simple proof of the number of regions. Their bijection was generalized to type C by Mészáros in [50]. In [7, Section 5.2] there is another proof of the formula for the number of regions, using Armstrong's and Rhoades' ceiling diagrams, which we define in Sect. 3.7. The ceiling diagrams are related to the diagrams Athanasiadis and Linusson used. Armstrong, Reiner, and Rhoades [6] define *nonnesting parking functions* using the root poset and permutations from the finite Weyl group and label the Shi regions with these. Their definition can also be used for types which are not crystallographic. Other, more recent and more general bijections include [16, 40] for example.

3.7 The Ish and the Shi

We'll start off by writing the q, t-Catalan polynomial combinatorially:

$$C_n(q,t) = \sum_{\pi} q^{\text{area}(\pi)} t^{\text{bounce}(\pi)}, \tag{3.3}$$

where the sum is over all Dyck paths of length n. The (q, t)-Catalan polynomials are remarkable generating functions coming from representation theory. They have been intensely studied since their introduction by Garsia and Haiman in [30]. See Haglund's monograph [35, Chapter 3] for more information. There are the same number of dominant Shi regions as there are Dyck paths, and one of Armstrong's results in [5] (and the one we'll describe) was to transfer the statistics area and bounce to dominant regions. His statistics are actually for all regions. The statistic shi will

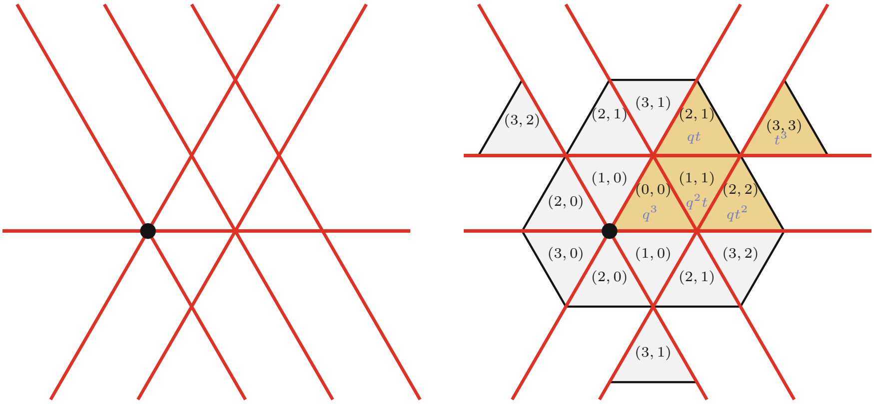


Fig. 9 The Ish arrangement for $n = 3$ is on the left. The Shi arrangement is on the right, where region R is labeled by the pair $(\mathrm{shi}(R), \mathrm{ish}(R))$. The dominant regions are also labeled by $q^{\binom{3}{2}-\mathrm{shi}(R)}t^{\mathrm{ish}(R)}$. The sum of the monomials is $C_3(q,t) = q^3 + q^2t + qt + qt^2 + t^3$

correspond to the statistic area and $\mathrm{shi}(R)$ is defined as the number of hyperplanes which must be crossed on a trip to the region R from $R_0 = A_0$. The statistic ish is defined using a second hyperplane arrangement, the Ish arrangement $\mathcal{I}sh_n$. It is defined for type A and is a deformation of the Coxeter arrangement. Let Δ be the set of roots for type A, so $\alpha_i = \varepsilon_i - \varepsilon_{i+1}$ as in Sect. 1.1.1. Denote $\alpha_i + \alpha_{i+1} + \cdots + \alpha_{n-1}$ by $\tilde{\alpha}_i$. Then the definition of the Ish arrangement is

$$\begin{aligned}\mathcal{I}sh_n &= \mathcal{C}ox_n \cup \{H_{\tilde{\alpha}_j,k} | 1 \le j \le n-1, k \in \{1, 2, \ldots, n-j\}\}\\ &= \mathcal{C}ox_n \cup \{x_j - x_n = k | k \in \{1, \ldots, n-j\}, 1 \le j \le n-1\}.\end{aligned}$$

The ish statistic is defined on Shi regions using the hyperplanes in $\mathcal{I}sh_n$. Let R be region of the arrangement $\mathcal{S}hi_n$ with minimal alcove A_R. There is a unique $w \in \widehat{\mathfrak{S}}_n$ such that $A_R = A_0 w$. The affine permutation w has a unique factorization $w^I \cdot w_I$ [17], where $w_I \in \mathfrak{S}_n$ and w^I is a *minimal length coset representative*, which we won't define. What is important for us is that $A_0 w^I$ is an alcove in the dominant chamber since w^I is a minimal length coset representative. Then $\mathrm{ish}(R) = \mathrm{ish}(A_R)$ is the number of hyperplanes in $\mathcal{I}sh_n$ which must be crossed in traveling from A_0 to Aw^I (Fig. 9).

Each dominant Shi region R corresponds to a Dyck path π_R. Armstrong showed that $\binom{n}{2} - \mathrm{shi}(R) = \mathrm{area}(\pi_R)$ and $\mathrm{ish}(R) = \mathrm{bounce}(\pi_R)$ Notice that the ish and shi statistics are defined on all regions, not just the dominant ones. Armstrong was able to show they agree with bounce and area on all *diagonally labeled* Dyck paths. See [35, Chapter 5] and [5, Section 3].

Armstrong and Rhoades concentrated on properties of the Ish arrangement, especially its uncanny similarities to the Shi arrangement, in [7]. Their definition of the arrangement changes just a bit: replace $x_j - x_n = i$ by $x_1 - x_{n-j+1} = n - i + 1$.

Their main theorem is for *deleted* versions (more general) of the arrangements (see Sect. 6), but we'll stick with the full arrangements. That is, G is the complete graph in this survey. We need to define a few terms before we can state the main theorem. The wall H of a region R is called a *ceiling* if it does not contain the origin and if the origin and R are not separated by H (they lie in the same half-space of H). The regions of both $\mathcal{I}sh_n$ and $\mathcal{S}hi_n$ are convex, so every region has a *recession cone*:

$$\mathcal{R}(R) = \{v \in V : v + R \subseteq R\}.$$

The cone is closed under nonnegative linear combinations and has a dimension. The dimension of $\mathcal{R}(R)$ is called the *degrees of freedom* of R. It's worth mentioning that the region R is bounded if and only if $\mathcal{R}(R) = \{0\}$.

A simplified version of their main theorem can now be stated:

Theorem 3.9 ([7])**.** Let c and d be nonnegative integers. The $\mathcal{I}sh_n$ and $\mathcal{S}hi_n$ have the same

(1) characteristic polynomial,
(2) number of dominant regions with c ceilings, and
(3) number of regions with c ceilings and d degrees of freedom.

We are not presenting their theorem in its full generality, and as written here, (1) was proved in [5].

We want to define the ceiling diagrams for the Shi arrangement because they show the properties of the corresponding region so clearly. We will also need variations on the root poset, which they use to prove Theorem 3.9. First the definition of the Shi ceiling diagram of a region R. Suppose the region is in the chamber Dw, where D is the dominant chamber and $w \in \mathfrak{S}_n$. Then we define the set partition σ_R: there is an arc from i to j, $i < j$, in the diagram of σ_R if and only if the hyperplane $x_{w(i)} - x_{w(j)} = 1$ is a ceiling of R. We draw the *Shi ceiling diagram* (w, σ_R) by placing the arc diagram for σ_R above $w(1), w(2), \ldots, w(n)$. See Example 3.10. Armstrong and Rhoades show that σ_R is a nonnesting set partition. The number of arcs is c. What about d? Let d' be the number of k, $1 \le k \le n-1$ where there is no arc covering the space between k and $k+1$; that is, the number of k where for which there is no $i < j$ such that $i \le k < j$ and there is an arc from i to j. For example, $d' = 0$ for the set partition in Fig. 2 and $d' = 1$ for the set partition 12|35|4. Then set $d = d' + 1$. Additionally, the recession cone $\mathcal{R}(R)$ can be read from the diagram.

There is still a key point: for a fixed $w \in \mathfrak{S}_n$, both the regions and the ceiling diagrams are in bijection with antichains in $\Delta^+(w)$. The poset $\Delta^+(w)$ is the first variation on the root poset:

$$\Delta^+(w) = \{x_{w(i)} - x_{w(j)} = 1 : w(i) < w(j)\}.$$

The elements of $\Delta^+(w)$ are the affine hyperplanes in the Shi arrangement which intersect Dw. The partial order on the hyperplanes is given by

$$x_{w(i')} - x_{w(j')} = 1 \le x_{w(i)} - x_{w(j)} = 1$$

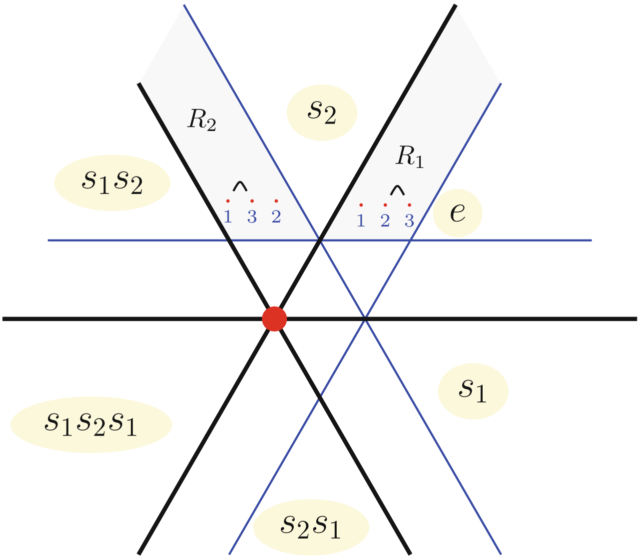


Fig. 10 The Shi arrangement for A_2. The chamber $(D)w$ is labeled by w

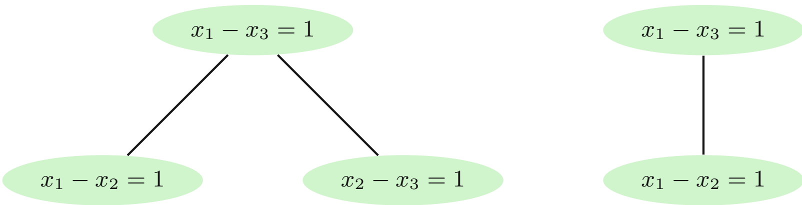


Fig. 11 On the left, the poset $\Delta^+(e)$. The poset $\Delta^+(s_2)$ is on the right

if $w(i) \leq w(i') < w(j') \leq w(j)$. The partial order is defined so that ceilings of any Shi region R, $R \in Dw$, are the maximal elements of an order ideal. The number c shows up as the number of these maximal elements.

We'll need the second variation on Δ^+, $\Psi^+(w)$, for discussing the Ish arrangement:

$$\Psi^+(w) = \{x_1 - x_j = i : w^{-1}(i) < w^{-1}(j)\}.$$

Its elements are the Ish hyperplanes that intersect wD and its partial order is chosen so that the ceilings of any Ish region R, $R \in wD$, are minimal elements of a filter, and c for the region is the number of these minimal elements.

Example 3.10. This example refers to Figs. 10 and 11. In Fig. 10, the chamber Dw is labeled by $w \in \mathfrak{S}_3$. We've picked two Shi regions to consider in this example-the ones we have labeled R_1 and R_2.

R_1 is in the chamber Dw for $w = e$, the identity. There are three hyperplanes of the form $x_i - x_j = 1$ which intersect Dw and the poset $\Delta^+(w)$ is on the left in Fig. 11. The hyperplane $x_2 - x_3 = 1$ is a ceiling for R_1 and the region is also labeled with its ceiling diagram in Fig. 10.

The region R_2 is in the chamber Dw for $w = s_2$ and poset $\Delta^+(s_2)$ is on the right in Fig. 11. The poset $\Delta^+(s_2)$ has three ideals, corresponding to the three Shi regions in Ds_2. Our region R_2 has ceiling $x_1 - x_3 = x_{w(1)} - x_{w(2)} = 1$ and we have placed the arc diagram for the partition $12|3$ above $w(1)w(2)w(3)$ to build the ceiling diagram.

The relationship between the filters and the regions is bijective in the Ish case, just as between ideals and regions in the Shi case. The posets $\Delta^+(w)$ and $\Psi^+(w)$ are dual to each other when w is the identity permutation e. The final step in the proof of Theorem 3.9, part (3), is simply to send an order ideal in $\Delta^+(e)$ to the corresponding filter in $\Psi^+(e)$. Since the maximal elements in the ideal become the minimal elements in the filter, c is preserved. There are also ceiling diagrams for the Ish arrangement, but we won't define them.

To prove Theorem 3.9, part (3), Armstrong and Rhoades used *ceiling partitions*, which are set partitions of $[n]$. We now define a simplified version of them. First suppose R is a Shi region. The ceiling partition π_R has an arc from i to j, $i < j$, if and only if the hyperplane $x_i - x_j = 1$ is a ceiling of R. Next suppose we have an Ish region R. Its ceiling partition has an arc from i to j, $i < j$, if and only if $x_1 - x_j = i$ is a ceiling of R. The definition of the ceiling partition does not depend on the chamber of R for either arrangement. Surprisingly, the distribution of the set partitions is the same for the Ish and Shi arrangements.

Theorem 3.11 ([7]). Let $\mathcal{A}$ be either the Ish or the Shi arrangement. Let π be a partition of $[n]$ with k blocks and let $1 \le d \le k$.

(1) The number of regions of $\mathcal{A}$ with ceiling partition π is

$$\frac{n!}{(n-k+1)!}.$$

(2) The number of regions of $\mathcal{A}$ with ceiling partition π and d degrees of freedom is

$$\frac{d(n-d-1)!(k-1)!}{(n-k-1)!(k-d)!}.$$

To obtain the number of regions with c ceilings and d degrees of freedom, thereby proving Theorem 3.9, part (3), sum the expression in Theorem 3.11, part (2), over all partitions π with $k = n - c$ blocks.

For space reasons, we cannot include the arguments here for Theorems 3.9 and 3.11. This is a shame, because we thereby don't present evidence for their observation [7]:

> The Ish arrangement is something of a "toy model" for the Shi arrangement (and other Catalan objects). That is, for any property P that $\mathcal{S}hi_n$ and $\mathcal{I}sh_n$ share, the proof that $\mathcal{I}sh_n$ satisfies P is easier than the proof that $\mathcal{S}hi_n$ satisfies P.

Many of the theorems of [7] are proved bijectively by Leven, Rhoades, and Wilson in [47].

3.8 Extended Shi Arrangement

In [54], Postnikov and Stanley introduced the extended Shi arrangement of type A_{n-1}:

$$\mathcal{S}hi_n^m = \{H_{\alpha,k} \mid \alpha \in \Delta^+, -m+1 \le k \le m\}.$$

This kind of extension is sometimes denoted by Fuss, as in *Fuss-Catalan* [4]. Up until now, we have been discussing $m = 1$. Postnikov and Stanley show that $\mathcal{S}hi_n^m$ has $(mn+1)^{n-1}$ regions. They fix m, set $f_n = \mathfrak{r}(\mathcal{S}hi_n^m)$ to be the number of regions, and show that the exponential generating function

$$f = \sum_{n \ge 0} f_n \frac{x^n}{n!}$$

satisfies

$$f = e^{xf^m}.$$

The extended Shi arrangement is a special case ($a = m, b = m+1$) of what they named *truncated affine arrangements*; see [54, Section 9] for more details. The dominant regions of the m-Shi are the same as the dominant regions of the m-Catalan.

Many of the other enumerative treats of the Shi arrangement generalize well to the extended Shi arrangement. For example, Stanley [68] labeled the m-Shi regions with m-parking functions of length n using an extended version of the bijection described here in Sect. 3.5. Stanley defined *m-parking functions of length n*. He replaced the condition that "$b_i \le i-1$" in the definition of parking function (see Sect. 3.5) by "$b_i \le m(i-1)$." If we set $d(R) = \text{area}(R)$ (see Sect. 3.7), then Stanley's bijection showed [68, Corollary 2.2] that

$$\sum_R q^{d(R)} = \sum_{(p_1,\ldots,p_n)} q^{p_1+\cdots+p_n}, \tag{3.4}$$

where the sum on the left is over all regions in $\mathcal{S}hi_n^m$ and the sum on the right is over all m-parking functions of length n (Fig. 12).

In 2004, Athanasiadis wrote two papers on the extended Catalan arrangement for crystallographic Δ, concentrating on the dominant regions. The Catalan arrangement has more hyperplanes than the Shi arrangement, but it has the same dominant regions, so we record his results here in terms of the m-Shi arrangement. We'll need a definition. The *Narayana numbers* (type A) are given by [53]

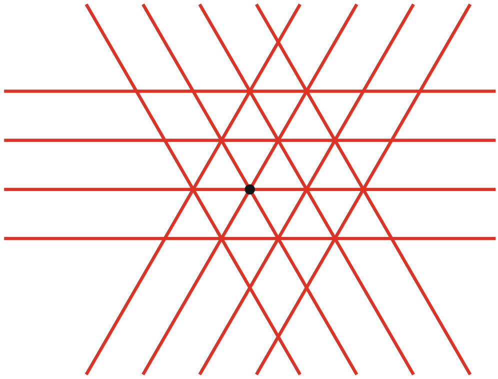


Fig. 12 The $\mathcal{S}hi_3^2$ arrangement. There are $49 = (2 \cdot 3 + 1)^2$ regions

$$N_{n,k} = \frac{1}{k+1}\binom{n}{k}\binom{n-1}{k}.$$

They refine the Catalan numbers by counting the Dyck paths of length n with k peaks. In other words, $C_n = \sum_{k=0}^{n-1} N_{n,k}$. Athanasiadis

(1) generalized and extended the Narayana numbers, finding what they enumerate in terms of dominant m-Shi regions;
(2) counted the number of m-Shi regions in the dominant chamber, generalizing Shi's result described in Sect. 3.1; and
(3) used co-filtered chains of ideals in the root poset to describe the dominant m-Shi regions.

We will describe (3) in a bit more depth. Let $\Delta^+ = I_0 \supseteq I_1 \supseteq I_2 \supseteq \cdots \supseteq I_m$ be a a decreasing chain $\mathcal{I}$ of ideals in Δ^+, set $I_i = I_m$ for all $i > m$, and set $J_i = \Delta^+ \setminus I_i$. The chain $\mathcal{I}$ is a *co-filtered chain of ideals of length m* if

(1) $(I_i + I_j) \cap \Delta^+ \subseteq I_{i+j}$ and
(2) $(J_i + J_j) \cap \Delta^+ \subseteq J_{i+j}$

is true for all indices $i, j \geq 1$ with $i + j \leq m$. The coordinates of the chain are

$$k_\alpha(\mathcal{I}) = \max\{k_1 + k_2 + \cdots + k - r : \alpha = \beta_1 + \cdots + \beta_r \text{ with } \beta_i \in I_{k_i} \text{ for all } i\}.$$

Athanasiadis showed that

$$k_\alpha(\mathcal{I}) + k_\beta(\mathcal{I}) \leq k_{\alpha+\beta}(\mathcal{I}) \leq k_\alpha(\mathcal{I}) + k_\beta(\mathcal{I}) + 1 \tag{3.5}$$

whenever $\alpha, \beta, \alpha + \beta \in \Delta^+$. Equation (3.5) generalizes Shi's bijection between filters in the root poset and dominant Shi regions (see Sect. 3.1). Finally, to define the fundamental m-Shi region associated to $\mathcal{I}$, he sets $R_\mathcal{I}$ to be the set of points $x \in V$ which satisfy

(1) $\langle \alpha \mid x \rangle > k$, if $\alpha \in I_k$ and
(2) $0 < \langle \alpha \mid x \rangle < k$, if $\alpha \in J_k$, for $0 \leq k \leq m$. The coordinates of the ideal are then the coordinates of a region.

Certain elements in an ideal are called indecomposable. These elements correspond to the walls of $R_{\mathcal{I}}$ which separate $R_{\mathcal{I}}$ from R_0, and take the place of peaks in Dyck paths when defining the Narayana numbers in terms of Shi regions.

Here we mention a few other enumerative results concerning the regions of the extended Shi arrangement. Any fixed hyperplane in the m-Shi arrangement is dissected into regions by the other hyperplanes in the arrangements. Fishel, Tzanaki, and Vazirani enumerate the number of regions for certain fixed hyperplanes in type A in [27]. Fishel, Kallipoliti, and Tzanaki [26] defined a bijection between dominant regions of the m-Shi arrangement in type A_n and dissections of an $m(n+1)+2$-gon. These dissections represent facets of the m-generalized cluster complex. In 2008, Sivasubramanian [65] gave combinatorial interpretations for the coeffiencients of a two-variable version of Stanley's distance enumerator (3.4) in type A for $m = 1$. Forge and Zaslavsky study the integral points in $[m]^m$ that do not lie in any hyperplane of the arrangement [29]. Thiel resolves a conjecture of Armstrong [4, Conjecture 5.1.24] on the distribution of floors and ceilings in the dominant regions of the m-Shi arrangements for all types. See Sect. 3.7 for the definition of ceiling.

4 Connections

4.1 Decompositions Numbers and the Shi Arrangement

The dominant regions make an appearance in the study of decomposition numbers for certain Hecke algebras. To describe this appearance, we'll first need a host of combinatorial definitions, then we'll indicate briefly how these arose from algebra, and finally we'll relate this back to the Shi arrangement. We thank Matthew Fayers for not only pointing out this connection, but carefully explaining it.

4.1.1 Combinatorics

Here we define n-cores, review some well-known facts about them, and review the abacus construction, which will be useful for us. Details can be found in [42].

The (k, l)-*hook* of an integer partition λ consists of the box in row k and column l of λ, all the boxes to the right of it in row k together with all the nodes below it and in column l. The *hook length* $h^{\lambda}_{(k,l)}$ of this box is the number of boxes in the (k, l)-hook. Let n be a positive integer. An *n-core* is a partition λ such that $n \nmid h^{\lambda}_{(k,l)}$ for all boxes λ. An *n-regular* partition has no (nonzero) n parts which equal each other. For example, $(7, 6, 6, 6)$ is not 3-regular. We'll sometimes use p or e instead of n, depending on the context. The definition is the same (Fig. 13).

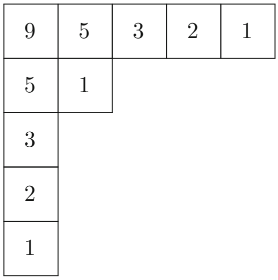


Fig. 13 The Young diagram of the partition $\lambda = (5, 2, 1, 1, 1)$. The hooklengths are the entries in the boxes of its Young diagram. The partition λ is a 4-core but not a 5-core

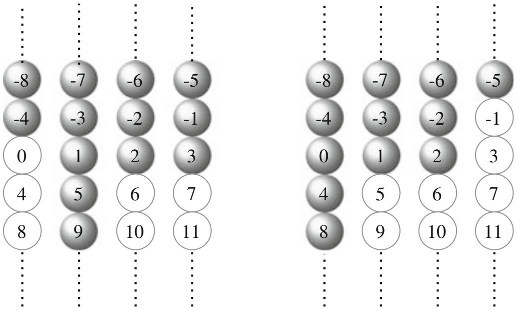


Fig. 14 On the left, the beads are placed on the positions labeled by the first column hook-lengths of $\lambda = (5, 2, 1, 1, 1)$. On the right is an equivalent abacus, where all bead positions have been shifted by $C = -1$

The β-*numbers* of the partition λ are the hook lengths from its first column. The β-numbers can be displayed on an abacus: a p-*abacus* is a diagram with p runners, labeled $0, 1, \ldots, p-1$. Runner i has positions labeled by integers $pj + i$, for all $j \in \mathbb{Z}$. We make a p-abacus for λ by placing a bead at position β_k, for each β-number β_k of λ and at all negative positions. We say two p-abaci are equivalent if we can change one to the other by moving the bead at position i to position $i + C$ for some $C \in \mathbb{Z}$ and for all positions i where there is a bead. The positive integer p is arbitrary for now, but will be related to the characteristic of a field when we see abaci in their algebraic context. See the 4-abacus in Fig. 14.

We can give an equivalent description of p-core partitions: a partition λ is an p-core if and only if whenever there is a bead at position j of its p-abacus, there is also a bead at position $j - p$ [42].

Suppose we have a partition which is not an p-core. Then there is at least one bead at a position j of its p-abacus which can be pushed up into the vacant position $j - p$. This gives us the p-abacus of another partition. We can repeat this until no beads can be pushed up, at which point we have the p-abacus of an p-core. The final partition γ is called the p-core of λ and the number of beads we moved in called the p-weight of λ. With a little work, which we won't do, it is possible to show that $|\lambda| = |\gamma| + wp$.

The last combinatorial ingredient we need are residues for the boxes of a partition. Let n be a positive integer. We call the box in row i, column j a k-*box* if $(j - i) \bmod n$ is equal to k.

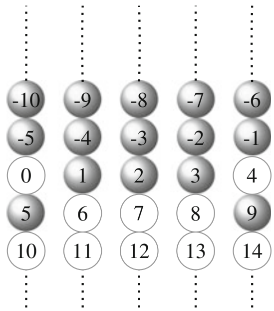


Fig. 15 The beads are placed on a 4-abacus on the positions labeled by the first column hook-lengths of $\lambda = (5, 2, 1, 1, 1)$. The partition λ is not a 5-core, since there are gaps on the 5-abacus. If we push the beads in positions 5 and 9 up to fill in the gaps, we obtain the 5-abacus of the empty partition

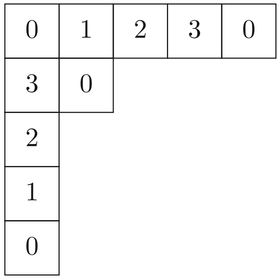


Fig. 16 The Young diagram of the partition $\lambda = (5, 2, 1, 1, 1)$. The entries are the residues modulo 4

4.1.2 Algebra

It is time to say a few words about how the combinatorics from the last section relate to algebra (Fig. 15).

Roughly speaking, the p-regular partitions of n index the irreducible modules D^λ of $F\mathcal{S}_n$, where p is the characteristic of the field F. The *p-blocks* are the equivalence classes of a certain equivalence relation on the irreducible modules. By Nakayama's celebrated conjecture, and Brauer and Robinson's theorem, two irreducibles D^λ and D^μ belong to the same p-block if and only λ and μ have the same p-core. Thus each p-block is labeled by a p-core and p-weight w. The p-weight keeps track of the difference between the p-core labeling a block and the partitions in the block: the p-core is a partition of $n - pw$. The weights are nonnegative integers. We will assume w is at least one, because the blocks where $w = 0$ are singletons consisting of the p-core (Fig. 16).

Scopes [57] investigated the classes of p-blocks under Morita equivalence. She characterized families of Morita equivalent p-blocks using the p-core and p-weight which label p-blocks. Suppose B is p-block for $F\mathfrak{S}_n$ labeled by weight w and p-core γ. Let k be an integer at least as large as w, and suppose that in a p-abacus for γ, there is a runner i which has k more beads in positive positions than runner $(i - 1)$ has. Now move k beads from runner i to runner $i - 1$ in such a p-abacus for γ. This new abacus determines another p-core, say $\bar{\gamma}$. See Fig. 17. The operation changes the size of the partition: $|\gamma| - k = |\bar{\gamma}|$. We are glossing over details here involving the β-numbers. Let $\bar{B}$ be the block of $F\mathfrak{S}_{n-k}$ labeled by w and $\bar{\gamma}$. Scopes took the

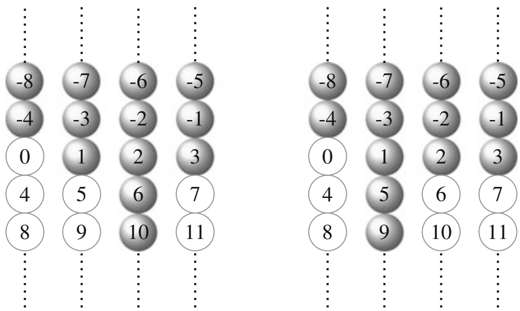


Fig. 17 On the left, the abacus for $\gamma = (6, 3, 1, 1, 1)$, on the right, the abacus for $\bar{\gamma} = (5, 2, 1, 1, 1)$. We have moved $k = 2$ beads from runner 2 to runner 1

transitive closure of the relation $B \sim \bar{B}$ and showed that within an equivalence class, the p-blocks have the same decomposition matrix, among other results.

Richards was studying the decomposition numbers for the Hecke algebra; see Sect. 2.1 for at least the definition. He used the classes from the Scopes equivalence on e-cores, where e depends on the characteristic p of the field and the element q used in the definition of the Hecke algebra. He called the classes *families*. Richards was interested in these families because the blocks of the Hecke algebras for $\mathfrak{S}_n$ and for $\mathfrak{S}_{n-k}$ corresponding to γ and $\bar{\gamma}$ respectively have essentially the same decomposition numbers [55].

Richards wanted to count such families. He built the following *pyramid* $\{{}_u a_v\}_{0 \leq u < v \leq e-1}$ for an e-core γ based on γ's e-abacus. Note the similarity in shape to admissible sign types and to the arrangement of roots in a staircase shape diagram in Fig. 2. For $i = 0, 1, \dots, e-1$, let p_i' be the position of the first free space on runner i. Arrange these e numbers in ascending order and relabel as $p_0 < p_1 < \cdots < p_{e-1}$. If $0 \leq u < v \leq e-1$, then $p_u - p_v$ is a positive integer not divisible by e. We may use any e-abacus for γ; it doesn't affect the set of differences. Richards defined the pyramid of numbers by

$$
{}_u a_v = \begin{cases} w-1 & \text{if } 0 < p_v - p_v < e \\ w-2 & \text{if } e < p_v - p_v < 2e \\ & \vdots \\ 1 & \text{if } (w-2)e < p_v - p_v < (w-1)e \\ 0 & \text{if } (w-1)e < p_v - p_v \end{cases}
$$

where w is the weight. Richards proved that two e-cores are in the same family if and only if they have the same pyramid and that there are exactly

$$
\frac{1}{e}\binom{ew}{e-1}
$$

families. What's more, he characterized the triangles of numbers which form a pyramid.

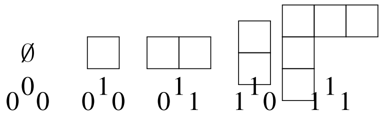
To show the connection to the Shi arrangement, we transform ${}_u a_v$ into ${}_u\tilde{a}_v$ by ${}_u a_v + {}_u\tilde{a}_v = w - 1$. Then Richard's Proposition 3.4 becomes

Proposition 4.1 ([55]). Let $e \geq 2$ and $w > 0$, and for $0 \leq u < v < e-1$ let $0 \leq {}_u\tilde{a}_v \leq w-1$. Then the ${}_u\tilde{a}_v$ form a pyramid if and only if for all $0 \leq u < t < v \leq e-1$,

$$
{}_u\tilde{a}_v = \begin{cases} {}_u\tilde{a}_t + {}_t\tilde{a}_v \text{ or } {}_u\tilde{a}_t + {}_t\tilde{a}_v + 1 & \text{if both of these have all entries no bigger than } w-1 \\ w-1 & \text{otherwise.} \end{cases}
$$

Please see Example 4.2 for the calculation of a few pyramids from cores.

Let's examine the case $e = 3$ and $w = 2$. There are five families. We choose five 3-cores $\{\emptyset, (1), (2), (1,1), (3,1,1)\}$ and calculate their pyramids:



The pyramids are all different, so we have found all the families. If we look back at the set G in Sect. 2.2 and consider the subset where all entries are either $+$ or $\bigcirc$, we see a similarity to the pyramids (replace $+$ with 1). This is true in general. Richard's proposition is the type A version of (3.5) from Sect. 3.8. The pyramid is also an admissible sign type for a dominant m-Shi region of type A, where $m = w - 1$.

4.1.3 Geometry

We'll just say a few more words about the geometry here. Richard's pyramids have connected the core partitions to regions. We'll describe Lascoux's [46] well-known bijection between n-cores and certain elements of $\widehat{\mathfrak{S}}_n$, and by extension, between n-cores and alcoves in the dominant chamber. Please see Lapointe's and Morse's paper [45] for details. We describe the bijection, as another way of seeing why core partitions pop up here. An n-core partition may have several removable boxes of a given residue or it may have several addable boxes of a given residue, but it will never have both addable and removable boxes of the same residue. Given an n-core partition λ and the generators $s_0, s_1, \ldots, s_{n-1}$ of $\widehat{\mathfrak{S}}_n$, let $s_i(\lambda)$ be the partition where all boxes of residue i have been removed (added) if there are removable (addable) boxes. Any n-core partition can be expressed as $w(\emptyset)$. See Fig. 18 and Example 4.2. We associate the n-core $w(\emptyset)$ with the alcove $A_0 w$.

We mention that Fishel and Vazirani mapped partitions which are both n and $nm+1$ cores to dominant regions in the m-Shi arrangment of type A_{n-1} in [28] using abacus diagrams and the root lattice.

Example 4.2. First, we construct the two pyramids for the partition $(4,2)$, one each for $w = 2$ and $w = 3$. From Fig. 19, we see that $(p'_0, p'_1, p'_2) = (p_0, p_1, p_2) =$

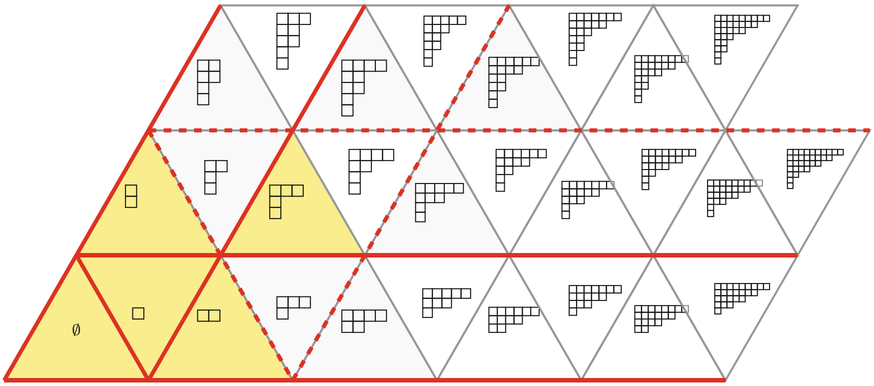
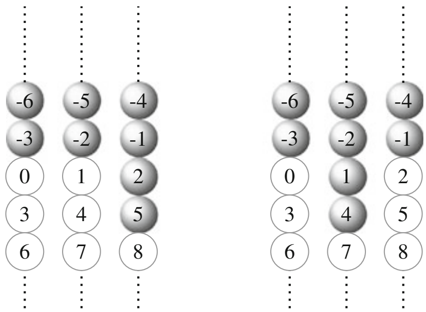


Fig. 18 The dominant chamber for A_2 with 3-cores. The thick solid lines are hyperplanes from the 1-Shi arrangement, the dashed ones are added to create the 2-Shi arrangement

Fig. 19 On the left, a 3-abacus diagram for (4, 2). On the right, a diagram for (3, 1). The right abacus is the result of moving two beads in the left abacus, as described in Sect. 4.1.2



$(0, 1, 8)$. When $w = 2$, the pyramid is $({}_0a_2, {}_0a_1, {}_1a_2) = (0, 1, 0)$ and when $w = 3$, it is $({}_0a_2, {}_0a_1, {}_1a_2) = (0, 2, 0)$. We also calculate the coordinates/admissible sign type of the region containing (4, 2)'s alcove. For $m = 1$ $(w = 2)$, $({}_0\tilde{a}_2, {}_0\tilde{a}_1, {}_1\tilde{a}_2) = (1, 0, 1)$ and for $m = 2$ $(w = 3)$, $({}_0\tilde{a}_2, {}_0\tilde{a}_1, {}_1\tilde{a}_2) = (2, 0, 2)$. Additionally,

$$s_0s_2s_1s_0(\emptyset) = s_0s_2s_1\left(\begin{array}{|c|}\hline 0\\\hline\end{array}\right) = s_0s_2\left(\begin{array}{|c|c|}\hline 0&1\\\hline\end{array}\right) = s_0\left(\begin{array}{|c|c|c|}\hline 0&1&2\\\hline 2\\\cline{1-1}\end{array}\right) = \begin{array}{|c|c|c|c|}\hline 0&1&2&0\\\hline 2&0\\\cline{1-2}\end{array},$$

where the entries in the boxes are their residues mod 3. We have placed (4, 2) in the alcove corresponding to $s_0s_2s_1s_0$. See Fig. 18.

Consider the 1-Shi region containing (4, 2)'s alcove. The minimal alcove of this region is labeled with the 3-core (2). To reach this region from A_0, we must cross only one translate each of $H_{\theta,1}$ and $H_{\alpha_2,0}$, and no translates of $H_{\alpha_1,0}$. This is reflected in ${}_0\tilde{a}_2 = 1$, ${}_1\tilde{a}_2 = 1$, and ${}_0\tilde{a}_1 = 0$ respectively. Next, we consider the 2-Shi region containing (4, 2)'s alcove. This region, whose minimal element alcove is labeled by (4, 2) itself, is separated from the fundamental alcove by two translates for each of $H_{\theta,1}$ and $H_{\alpha_2,0}$, and no translates of $H_{\alpha_1,0}$, reflected in the $m = 2$ pyramid for (4, 2).

We repeat the calculations for the 3-core (3, 1), whose 3-abacus is on the right in Fig. 19. We calculate

$$s_2s_1s_0(\emptyset) = s_2s_1(\begin{array}{|c|}\hline 0\\\hline\end{array}) = s_2(\begin{array}{|c|c|}\hline 0&1\\\hline\end{array}) = \begin{array}{|c|c|c|}\hline 0&1&2\\\hline 2\\\cline{1-1}\end{array},$$

so we have placed (3, 1) in the alcove corresponding to $s_2s_1s_0$. The $w = 2$ pyramid for (3, 1) is $({}_0a_2, {}_0a_1, {}_1a_2) = (0, 1, 0)$, the same as the pyramid for (4, 2), indicating that its alcove will be in the same 1-Shi region as the alcove labeled by (4, 2). Its $w = 3$ pyramid is $({}_0a_2, {}_0a_1, {}_1a_2) = (0, 2, 1)$, which is not the same as the $w = 3$ pyramid for (4, 2), and indeed, their alcoves are in different 2-Shi regions.

Lastly, we mention that the action of moving beads as described in Sect. 4.1.2 corresponds to flipping the alcove containing the core of the original abacus over a hyperplane, to the alcove containing the core obtained through the bead move.

4.2 Finite Automata and Reduced Expressions

Headley used the Shi arrangement to build an automaton which recognizes reduced expressions. A *language* $\mathcal{L}$ is a subset of the set B^* of words in a given finite alphabet B. For us, the alphabet will be the set S of a Coxeter group W and the language will be reduced expressions.

A *finite state automaton* is a finite directed graph, with one vertex designated as the *initial state* S_0 and a subset of vertices as final states, and with every edge labeled by an element of B. We call the vertices states. A word $w \in \mathcal{L}$ is accepted by the automaton if the sequence of edge labels along some directed path starting at S_0 and ending at a final state is equal to w.

Headley was not the first nor the last to construct an automaton to accept reduced words; see Björner and Brenti [17], Hohlweg, Nadeau, and Williams [39], and Gunnells [33], for instance. However, Headley realized that if $s_{i_2}s_{i_3}\cdots s_{i_k}$ is reduced, then the fundamental alcove A_0 and $A_0s_{i_2}s_{i_3}\cdots s_{i_k}$ lie on the same side of the hyperplane fixed by s_{i_1} if and only if $s_{i_1}s_{i_2}s_{i_3}\cdots s_{i_k}$ is reduced. He built his automaton on this observation. The key lemma is

Lemma 4.3. [37] Let W be an irreducible affine Weyl group with root system Δ. Let R be a region of the Shi arrangement. If R and A_0 lie on the same side of the hyperplanes fixed by $s \in S$, then Rs lies in a single region.

The states of his automaton are the regions of the Shi arrangement. The fundamental alcove A_0, which is also a Shi region, is the initial state. All states are final. Let $s \in S$, and let R be a region. If R and A_0 are on the same side of the hyperplane fixed by $s \in S$, then let R' be the region containing Rs and place an arrow from R to R', labeled by s. If R and A_0 are not on the same side of the hyperplane, then R has no outgoing arrow labeled by s. Headley not only showed that the language accepted by this automaton is the set of all reduced words, he showed that if W is

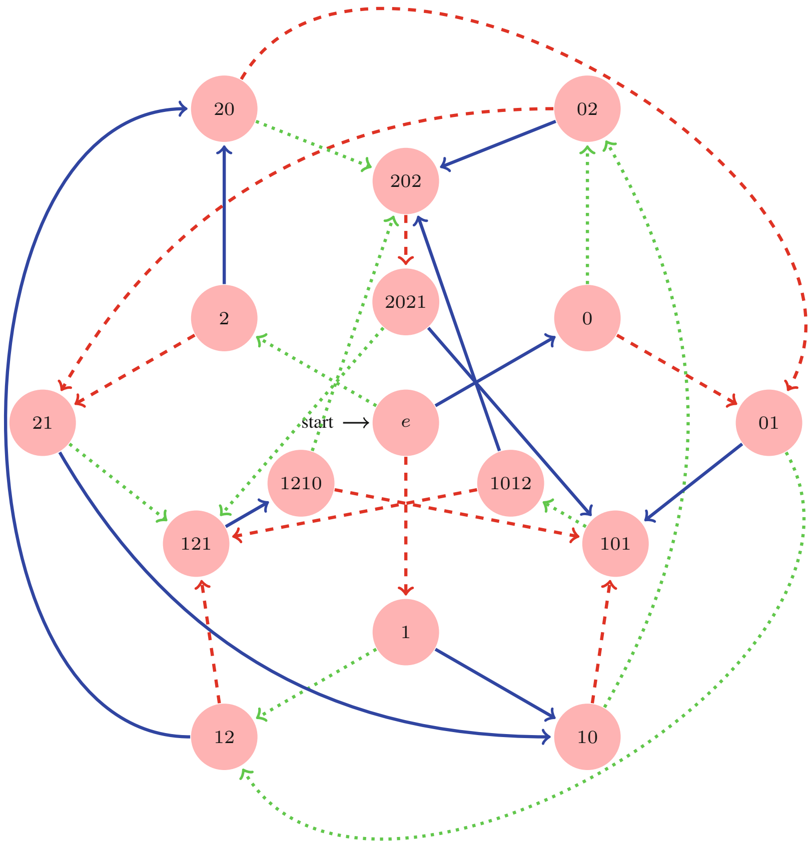


Fig. 20 Headley's automaton based on the Shi arrangement. It accepts reduced words in $\widehat{\mathfrak{S}}_3$. Each state is labeled with the affine permutation corresponding to the minimal alcove of the region. We use i instead of s_i. Solid arrows represent an edge labeled by s_0 and dashed (respectively dotted) edges represent edges labeled s_1 (respectively s_2). See Example 4.4

the affine symmetric group, the automaton has the minimal number of states. There were actually two automata in Headley's thesis. The first produced a nice generating function, but it has more states.

Example 4.4. This example refers to Fig. 20. The path

$$e \xrightarrow{0} 0 \xrightarrow{2} 02 \xrightarrow{1} 21 \xrightarrow{0} 10 \xrightarrow{1} 101 \xrightarrow{2} 1012$$

represents the expression $s_0s_2s_1s_0s_1s_2$, which is reduced. Since $s_0s_1s_0s_1$ is not reduced, there is no path which starts at e and follows solid, dashed, solid, dashed arrows.

4.3 More Connections

This connection is to the filters in Δ^+, not to the Shi arrangement directly. In [20–22], Cellini and Papi investigate ad-nilpotent ideals in a Borel subalgebra. They associate each ideal to a filter in Δ^+ and also to an element of the affine Weyl group. Very roughly speaking, they use the Cartan decomposition $L = H \oplus N$ and $N = \bigoplus_{\alpha\in\Delta^+} L_\alpha$ and the definition of an ad-nilpotent ideal as an ideal contained in N to define the antichain

$$\Delta_I = \{\alpha \in \Delta^+ : L_\alpha \subseteq I\}$$

which defines a filter. See also Suter [73]. Dong extends Cellini and Papi's work from Borel subalgebras to parabolic subalgebras in [24]. He uses deleted Shi arrangements, which we don't address in this survey. Panyushev [52] developes combinatorial aspects of the theory of ad-nilpotent ideals, giving a geometric interpretationfor the number of generators of an ideal, for example.

Gunnells and Sommers study *Dynkin elements*, which we won't define, in [34]. They define N-regions, which turn out to be unions of Shi regions. A simplified version of their main theorem is that if x is the point of minimal Euclidean length in the closure of an N-region, then $2x$ is a Dynkin element.

5 Further Developments

We briefly mention a few recent results. In [31], Gorsky, Mazin, and Vazirani developed "rational slope" versions of much of what has been discussed here. A tuple $(b_1, \ldots, b_n)$ of nonnegative integers is called an M/n-parking function if the Young diagram with row lengths equal to $b_1, \ldots, b_n$ arranged in decreasing order fits above the diagonal in an $n \times M$ rectangle. We've stated it a bit differently than in Sects. 3.5 and 3.8, but if we let $M = n + 1$ and $M = mn + 1$ respectively and reverse the order of the tuple, we obtain the same functions. Gorsky, Mazin, and Vazirani defined M-*stable* permutations to take the place of minimal permutations of Shi regions and generalized the Pak–Stanley bijection, as well as another map defined by Anderson [2] in her study of core partitions. They conjectured their generalization of the Pak–Stanley map is injective for all relatively prime M and n. In 2017, McCammond, Thomas, and Williams [49] proved the conjectures in [31]. Additionally, Gorsky, Mazin, and Vazirani connect their maps to the combinatorics of q, t-Catalan polynomials. Sulzgruber [72] built on [31] by finding the coordinates of the M-stable permutations, generalizing (3.5). Thiel [74] extended their work to other types, among other results.

As mentioned in Sect. 3 Hohlweg, Nadeau, and Williams generalized the Shi arrangement to any Coxeter group, using n-small roots, and then to indefinite Coxeter systems. They also investigated automata.

6 Themes We Haven't Included

We give a short and incomplete list of topics we have not discussed.

(1) The Shi arrangement is free. Either see original article by Athanasiadis [9] or his excellent summary [10]. Abe, Suyama, and Tsujie [1] show that the Ish arrangement is free.
(2) In graphical arrangements or deleted arrangements, some of the hyperplanes have been removed. We survey only the complete Shi arrangement.
(3) We have no discussion of the connections to the torus $\check{Q}/(1+mh)\check{Q}$, where $\check{Q}$ is the coroot lattice of a root system, $(mh+1)\check{Q}$ is its dilate, and h is the Coxeter number of the root system. See Athanasiadis [12] or Haiman [36] for more information.
(4) The enumeration of bounded regions has nice results, which we have not discussed. See Athanasiadis and Tzanaki [15], for example and Sommers [66].

Acknowledgements The author would like to thank Hélène Barcelo, Gizem Karaali, and Rosa Orellana for allowing her to write an article for this AWM series. She would like to thank Matthew Fayers, Sarah Mason, and Jian-Yi Shi for comments on the manuscript, and Nathan Williams and Brendon Rhoades for suggesting several related papers. She would like to thank Patrick Headley for help with his thesis. The anonymous referees' comments helped enormously to improve exposition. This work was supported by a grant from the Simons Foundation (#359602, Susanna Fishel).

References

1. T. Abe, D. Suyama, S. Tsujie, The freeness of Ish arrangements. J. Comb. Theory Ser. A **146**, 169–183 (2017). MR 3574228
2. J. Anderson, Partitions which are simultaneously t_1- and t_2-core. Discret. Math. **248**(1–3), 237–243 (2002). MR 1892698
3. F. Ardila, Computing the Tutte polynomial of a hyperplane arrangement. Pac. J. Math. **230**(1), 1–26 (2007). MR 2318445
4. D. Armstrong, *Generalized Noncrossing Partitions and Combinatorics of Coxeter Groups*, Memoirs of the American Mathematical Society (2009), x+159 p. MR 2561274
5. D. Armstrong, Hyperplane arrangements and diagonal harmonics. J. Comb. **4**(2), 157–190 (2013). MR 3096132
6. D. Armstrong, V. Reiner, B. Rhoades, Parking spaces. Adv. Math. **269**, 647–706 (2015). MR 3281144
7. D. Armstrong, B. Rhoades, The Shi arrangement and the Ish arrangement. Trans. Am. Math. Soc. **364**(3), 1509–1528 (2012). MR 2869184
8. C.A. Athanasiadis, Characteristic polynomials of subspace arrangements and finite fields. Adv. Math. **122**(2), 193–233 (1996). MR 1409420
9. C.A. Athanasiadis, On free deformations of the braid arrangement. Eur. J. Comb. **19**(1), 7–18 (1998). MR 1600259
10. C.A. Athanasiadis, Deformations of Coxeter hyperplane arrangements and their characteristic polynomials, *Arrangements-Tokyo 1998*. Advanced Studies in Pure Mathematics, vol. 27 (Kinokuniya, Tokyo, 2000), pp. 1–26. MR 1796891
11. C.A. Athanasiadis, Generalized Catalan numbers, Weyl groups and arrangements of hyperplanes. Bull. Lond. Math. Soc. **36**(3), 294–302 (2004). MR 2038717

12. C.A. Athanasiadis, On a refinement of the generalized Catalan numbers for Weyl groups. Trans. Am. Math. Soc. **357**(1), 179–196 (2005). MR 2098091
13. C.A. Athanasiadis, A combinatorial reciprocity theorem for hyperplane arrangements. Can. Math. Bull. **53**(1), 3–10 (2010). MR 2583206
14. C.A. Athanasiadis, S. Linusson, A simple bijection for the regions of the Shi arrangement of hyperplanes. Discret. Math. **204**(1–3), 27–39 (1999). MR 1691861
15. C.A. Athanasiadis, E. Tzanaki, On the enumeration of positive cells in generalized cluster complexes and Catalan hyperplane arrangements. J. Algebra. Comb. **23**(4), 355–375 (2006). MR 2236611
16. M. Beck, A. Berrizbeitia, M. Dairyko, C. Rodriguez, A. Ruiz, S. Veeneman, Parking functions, Shi arrangements, and mixed graphs. Am. Math. Mon. **122**(7), 660–673 (2015). MR 3383893
17. A. Björner, F. Brenti, Combinatorics of Coxeter groups, *Graduate Texts in Mathematics*, vol. 231 (Springer, New York, 2005). MR 2133266
18. A. Blass, B.E. Sagan, Characteristic and Ehrhart polynomials. J. Algebra. Comb. **7**(2), 115–126 (1998). MR 1609889
19. N. Bourbaki, éléments de mathématique. Fasc. XXXIV. Groupes et algèbres de Lie. Chapitre IV: Groupes de Coxeter et systèmes de Tits. Chapitre V: Groupes engendrés par des réflexions. Chapitre VI: systèmes de racines, *Actualités Scientifiques et Industrielles*, vol. 1337 (Hermann, Paris, 1968). MR 0240238
20. P. Cellini, P. Papi, Ad-nilpotent ideals of a Borel subalgebra. J. Algebra **225**(1), 130–141 (2000). MR 1743654
21. P. Cellini, P. Papi, Ad-nilpotent ideals of a Borel subalgebra II. J. Algebra **258**(1), 112–121 (2002). Special issue in celebration of Claudio Procesi's 60th birthday. MR 1958899
22. P. Cellini, P. Papi, Abelian ideals of Borel subalgebras and affine Weyl groups. Adv. Math. **187**(2), 320–361 (2004). MR 2078340
23. H.H. Crapo, G-C. Rota, *On the Foundations of Combinatorial Theory: Combinatorial Geometries*, preliminary ed., (MIT Press, Cambridge, London, 1970). MR 0290980
24. C.-P. Dong, Ad-nilpotent ideals and the Shi arrangement. J. Comb. Theory Ser. A **120**(8), 2118–2136 (2013). MR 3102177
25. R. Duarte, A.G. de Oliveira, The braid and the Shi arrangements and the Pak-Stanley labelling. Eur. J. Comb. **50**, 72–86 (2015). MR 3361413
26. S. Fishel, M. Kallipoliti, E. Tzanaki, Facets of the generalized cluster complex and regions in the extended Catalan arrangement of type A. Electron. J. Comb. **20**(4), 7–21 (2013). MR 3139392
27. S. Fishel, E. Tzanaki, M. Vazirani, Counting Shi regions with a fixed separating wall. Ann. Comb. **17**(4), 671–693 (2013). MR 3129778
28. S. Fishel, M. Vazirani, A bijection between dominant Shi regions and core partitions. Eur. J. Comb. **31**(8), 2087–2101 (2010). MR 2718283
29. D. Forge, T. Zaslavsky, Lattice point counts for the Shi arrangement and other affinographic hyperplane arrangements. J. Comb. Theory Ser. A **114**(1), 97–109 (2007). MR 2275583
30. A.M. Garsia, M. Haiman, A remarkable q, t-Catalan sequence and q-Lagrange inversion. J. Algebra. Comb. **5**(3), 191–244 (1996). MR 1394305
31. E. Gorsky, M. Mazin, M. Vazirani, Affine permutations and rational slope parking functions. Trans. Am. Math. Soc. **368**(12), 8403–8445 (2016). MR 3551576
32. P.E. Gunnells, Cells in Coxeter groups. Not. Am. Math. Soc. **53**(5), 528–535 (2006). MR 2254399
33. P.E. Gunnells, Automata and cells in affine Weyl groups. Represent. Theory **14**, 627–644 (2010). MR 2726285
34. P.E. Gunnells, E. Sommers, A characterization of Dynkin elements. Math. Res. Lett. **10**(2–3), 363–373 (2003). MR 1981909
35. J. Haglund, *The q,t-Catalan Numbers and the Space of Diagonal Harmonics*. University Lecture Series, vol. 41 (American Mathematical Society, Providence, 2008). With an appendix on the combinatorics of Macdonald polynomials. MR 2371044

36. M.D. Haiman, Conjectures on the quotient ring by diagonal invariants. J. Algebra. Comb. **3**(1), 17–76 (1994). MR 1256101
37. P.T. Headley, Reduced expressions in infinite Coxeter groups, Ph.D. Thesis, University of Michigan (ProQuest LLC, Ann Arbor, MI, 1994). MR 2691313
38. P. Headley, On a family of hyperplane arrangements related to the affine Weyl groups. J. Algebra. Comb. **6**(4), 331–338 (1997). MR 1471893
39. C. Hohlweg, P. Nadeau, N. Williams, Automata, reduced words and Garside shadows in Coxeter groups. J. Algebra **457**, 431–456 (2016). MR 3490088
40. S. Hopkins, D. Perkinson, Bigraphical arrangements. Trans. Am. Math. Soc. **368**(1), 709–725 (2016). MR 3413881
41. J.E. Humphreys, *Reflection Groups and Coxeter Groups*. Cambridge Studies in Advanced Mathematics, vol. 29 (Cambridge University Press, Cambridge, 1990). MR 1066460
42. G. James, A. Kerber, *The Representation Theory of the Symmetric Group*. Encyclopedia of Mathematics and its Applications, vol. 16 (Addison-Wesley Publishing Co., Reading, 1981). With a foreword by P. M. Cohn, With an introduction by Gilbert de B. Robinson. MR 644144
43. V.G. Kac, *Infinite-dimensional Lie Algebras*, 3rd edn. (Cambridge University Press, Cambridge, 1990). MR 1104219
44. D. Kazhdan, G. Lusztig, Representations of Coxeter groups and Hecke algebras. Invent. Math. **53**(2), 165–184 (1979). MR 560412
45. L. Lapointe, J. Morse, Tableaux on $k+1$-cores, reduced words for affine permutations, and k-Schur expansions. J. Comb. Theory Ser. A **112**(1), 44–81 (2005). MR 2167475
46. A. Lascoux, Ordering the affine symmetric group, *Algebraic Combinatorics and Applications*, Gößweinstein, 1999 (Springer, Berlin, 2001), pp. 219–231. MR 1851953
47. E. Leven, B. Rhoades, A.T. Wilson, Bijections for the Shi and Ish arrangements. Eur. J. Comb. **39**, 1–23 (2014). MR 3168512
48. G. Lusztig, Some examples of square integrable representations of semisimple p-adic groups. Trans. Am. Math. Soc. **277**(2), 623–653 (1983). MR 694380
49. J. McCammond, H. Thomas, N. Williams, *Fixed points of parking functions* (2017)
50. K. Mészáros, Labeling the regions of the type C_n Shi arrangement. Electron. J. Comb. **20**(2), Paper 31, 12 (2013). MR 3066370
51. J.W. Moon, in *Counting Labelled Trees*, From lectures delivered to the Twelfth Biennial Seminar of the Canadian Mathematical Congress, Vancouver, vol. 1969 (Que, Canadian Mathematical Congress, Montreal, 1970). MR 0274333
52. D.I. Panyushev, Ad-nilpotent ideals of a Borel subalgebra: generators and duality. J. Algebra **274**(2), 822–846 (2004). MR 2043377
53. T. Kyle Petersen, *Eulerian Numbers*, Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks] (Birkhäuser/Springer, New York, 2015). With a foreword by Richard Stanley. MR 3408615
54. A. Postnikov, Deformations of Coxeter hyperplane arrangements. J. Comb. Theory Ser. A **91**(1–2), 544–597 (2000). In memory of Gian-Carlo Rota. MR 1780038
55. M.J. Richards, Some decomposition numbers for Hecke algebras of general linear groups. Math. Proc. Camb. Philos. Soc. **119**(3), 383–402 (1996). MR 1357053
56. F. Rincón, A labelling of the faces in the Shi arrangement, Rose-Hulman Undergrad. Math. J. **8**(1), 7 (2007)
57. J. Scopes, Cartan matrices and Morita equivalence for blocks of the symmetric groups. J. Algebra **142**(2), 441–455 (1991). MR 1127075
58. J.Y. Shi, *The Kazhdan-Lusztig Cells in Certain Affine Weyl Groups*. Lecture Notes in Mathematics, vol. 1179 (Springer, Berlin, 1986). MR 835214
59. J.Y. Shi, Alcoves corresponding to an affine Weyl group. J. Lond. Math. Soc. (2) **35**(1), 42–55 (1987). MR 871764
60. J.Y. Shi, Sign types corresponding to an affine Weyl group. J. Lond. Math. Soc. (2) **35**(1), 56–74 (1987). MR 871764
61. J.Y. Shi, A two-sided cell in an affine Weyl group. J. Lond. Math. Soc. (2) **36**(3), 407–420 (1987). MR 918633

62. J.Y. Shi, A two-sided cell in an affine Weyl group II. J. Lond. Math. Soc. (2) **37**(2), 253–264 (1988). MR 918633
63. J.Y. Shi, Jian Yi Shi, Left cells in the affine Weyl group $W_a(\widetilde{D}_4)$. Osaka J. Math. **31**(1), 27–50 (1994). MR 1262787
64. J.-Y. Shi, The number of $\oplus$-sign types. Q. J. Math. Oxf. Ser. (2) **48**(189), 93–105 (1997). MR 1439701
65. S. Sivasubramanian, On the two variable distance enumerator of the Shi hyperplane arrangement. Eur. J. Comb. **29**(5), 1104–1111 (2008). MR 2419213
66. N. Eric, Sommers, B-stable ideals in the nilradical of a Borel subalgebra. Can. Math. Bull. **48**(3), 460–472 (2005). MR 2154088
67. P. Richard, Stanley, Hyperplane arrangements, interval orders, and trees. Proc. Natl. Acad. Sci. USA **93**(6), 2620–2625 (1996). MR 1379568
68. R.P. Stanley, Hyperplane arrangements, parking functions and tree inversions, *Mathematical Essays in Honor of Gian-Carlo Rota (Cambridge, MA, 1996)*. Progress in Mathematics, vol. 161 (Birkhäuser, Boston, 1998), pp. 359–375. MR 1627378
69. R.P. Stanley, An introduction to hyperplane arrangements, *Geometric Combinatorics*, IAS/Park City Mathematics Series (American Mathematical Society, Providence, 2007), pp. 389–496. MR 2383131
70. R.P. Stanley, *Enumerative Combinatorics, Volume 1*, 2nd edn. Cambridge Studies in Advanced Mathematics, vol. 49 (Cambridge University Press, Cambridge, 2012). MR 2868112
71. R.P. Stanley, *Catalan Numbers* (Cambridge University Press, New York, 2015). MR 3467982
72. R. Sulzgruber, Rational Shi tableaux and the skew length statistic (2015), arXiv:1512.04320
73. R. Suter, Abelian ideals in a Borel subalgebra of a complex simple Lie algebra. Invent. Math. **156**(1), 175–221 (2004). MR 2047661
74. M. Thiel, On floors and ceilings of the k-Catalan arrangement. Electron. J. Comb. **21**(4), Paper 4.36, 15. MR 3292273
75. H. Thomas, N. Williams, Cyclic symmetry of the scaled simplex. J. Algebra. Comb. **39**(2), 225–246 (2014). MR 3159251
76. T. Zaslavsky, Counting the faces of cut-up spaces. Bull. Am. Math. Soc. **81**(5), 916–918 (1975). MR 0400066
77. Thomas Zaslavsky, Facing up to arrangements: face-count formulas for partitions of space by hyperplanes. Memoirs Am. Math. Soc. **1**(154), vii+102 p. (1975). MR 0357135

Variations on a Theme of Schubert Calculus

Maria Gillespie

Abstract In this tutorial, we provide an overview of many of the established combinatorial and algebraic tools of Schubert calculus, the modern area of enumerative geometry that encapsulates a wide variety of topics involving intersections of linear spaces. It is intended as a guide for readers with a combinatorial bent to understand and appreciate the geometric and topological aspects of Schubert calculus, and conversely for geometric-minded readers to gain familiarity with the relevant combinatorial tools in this area. We lead the reader through a tour of three variations on a theme: Grassmannians, flag varieties, and orthogonal Grassmannians. The orthogonal Grassmannian, unlike the ordinary Grassmannian and the flag variety, has not yet been addressed very often in textbooks, so this presentation may be helpful as an introduction to type B Schubert calculus. This work is adapted from the author's lecture notes for a graduate workshop during the Equivariant Combinatorics Workshop at the Center for Mathematics Research, Montreal, June 12–16, 2017.

1 Introduction

Schubert calculus was invented as a general method for solving linear intersection problems in Euclidean space. One very simple example of a linear intersection problem is the following: How many lines pass through two given points in the plane?

It is almost axiomatically true that the answer is 1, as long as the points are distinct (otherwise it is ∞). Likewise, we can ask how many points are contained in two lines in the plane. The answer is also usually 1, though it can be 0 if the lines are parallel, or ∞ if the lines are equal.

In higher dimensions, the answers may change: in three-dimensional space, there are most often zero points of intersection of two given lines. One can also consider

Supported by NSF MSPRF grant PDRF 1604262.

M. Gillespie (✉)
University of California, Davis, CA, USA
e-mail: mgillespie@math.ucdavis.edu

H. Barcelo et al. (eds.), *Recent Trends in Algebraic Combinatorics*,
Association for Women in Mathematics Series 16,
https://doi.org/10.1007/978-3-030-05141-9_4

more complicated intersection problems involving subspaces of Euclidean space. For instance, how many planes in 4-space contain a given line and a given point? Generically, the answer will be 1, but in degenerate cases (when the point is on the line) may be ∞.

It seems that the answers to such problems are often 1, 0, or ∞, but this is not always the case. Here is the classic example of Schubert calculus, where the answer is generically 2:

Question 1.1. *How many lines intersect four given lines in three-dimensional space?*

Hermann Schubert's nineteenth-century solution to this question[1] would have invoked what he called the "Principle of Conservation of Number" as follows. Suppose the four lines l_1, l_2, l_3, l_4 were arranged so that l_1 and l_2 intersect at a point P, l_2 and l_3 intersect at Q and none of the other pairs of lines intersect and the planes ρ_1 and ρ_2 determined by l_1, l_2 and l_3, l_4, respectively, are not parallel. Then ρ_1 and ρ_2 intersect at another line α, which necessarily intersects all four lines. The line β through P and Q also intersects all four lines, and it is not hard to see that these are the only two in this case.

Schubert would have said that since there are two solutions in this configuration, there are two for every configuration of lines for which the number of solutions is finite, since the solutions can be interpreted as solutions to polynomial equations over the complex numbers. The answer is indeed preserved in this case, but the lack of rigor in this method regarding multiplicities led to some errors in computations in harder questions of enumerative geometry.

The following is an example of a more complicated enumerative geometry problem, which is less approachable with elementary methods.

Question 1.2. *How many k-dimensional subspaces of $\mathbb{C}^n$ intersect each of $k \cdot (n-k)$ fixed subspaces of dimension $n-k$ nontrivially?*

Hilbert's 15th problem asked to put Schubert's enumerative methods on a rigorous foundation. This led to the modern day theory known as Schubert calculus.

The main idea, going back to Question 1.1, is to let X_i be the space of all lines L intersecting l_i for each $i = 1, \ldots, 4$. Then the intersection $X_1 \cap X_2 \cap X_3 \cap X_4$ is the set of solutions to our problem. Each X_i is an example of a *Schubert variety*, an algebraic and geometric object that is essential to solving these types of intersection problems.

[1] See [49] for Schubert's original work, or [45] for a modern exposition on Schubert's methods.

1.1 "Variations on a Theme"

This tutorial on Schubert calculus is organized as a theme and variations.[2] In particular, after briefly recalling some of the necessary geometric background on projective spaces in Sect. 2 (which may be skipped or skimmed over by readers already familiar with these basics), we begin in Sect. 3 (the 'Theme') with the foundational ideas of Schubert calculus going back to Schubert [49]. This includes a rigorous development of Schubert varieties in the *Grassmannian*, the set of all k-dimensional subspaces of a fixed n-dimensional space, and a more careful geometric analysis of the elementary intersection problems mentioned in the introduction. We also establish the basic properties of the Grassmannian. Much of this background material can also be found in expository sources such as [22, 28, 39], and much of the material in the first few sections is drawn from these works.

In Variation 1 (Sect. 4), we present the general formulas for intersecting complex Schubert varieties and show how it relates to calculations in the cohomology of the Grassmannian as well as products of Schur functions. Variation 2 (Sect. 5) repeats this story for the complete flag variety (in place of the Grassmannian), with the role of Schur functions replaced by the Schubert polynomials. Finally, Variation 3 (Sect. 6) explores Schubert calculus in the "Lie type B" Grassmannian, known as the *orthogonal Grassmannian*.

There are countless more known variations on the theme of classical Schubert calculus, including Grassmannians in the remaining Lie types, partial flag varieties, and Schubert varieties over the real numbers. There is also much that has yet to be explored. We conclude with an overview of some of these potential further directions of study in Sect. 7.

2 Background on Projective Space

The notion of *projective space* helps clean up many of the ambiguities in the question above. For instance, in the projective plane, parallel lines meet, at a "point at infinity".[3] It also is one of the simplest examples of a Schubert variety (Fig. 1).

One way to define projective space over a field k is as the set of lines through the origin in one higher dimensional space as follows.

Definition 2.1. The n-dimensional **projective space** $\mathbb{P}^n_k$ over a field k is the set of equivalence classes in $k^{n+1} \setminus \{(0, 0, \ldots, 0)\}$ with respect to the relation $\sim$ given by scalar multiplication, that is,

$$(x_0, x_1, \ldots, x_n) \sim (y_0, y_1, \ldots, y_n)$$

[2] A play on words that references the shared surname with musical composer Franz Schubert, who also lived in Germany in the nineteenth century.

[3] Photograph of the train tracks downloaded from edupic.net.

Fig. 1 Parallel lines meeting at a point at infinity

if and only if there exists $a \in k \setminus \{0\}$ such that $ax_i = y_i$ for all i. We write $(x_0 : x_1 : \cdots : x_n)$ for the equivalence class in $\mathbb{P}^k$ containing $(x_0, \ldots, x_n)$, and we refer to $(x_0 : x_1 : \cdots : x_n)$ as a *point* in $\mathbb{P}^k$ in *homogeneous coordinates*.

Note that *a point in* $\mathbb{P}^n_k$ *is a line through the origin in* k^{n+1}. In particular, a line through the origin consists of all scalar multiples of a given nonzero vector.

Unless we specify otherwise, we will always use $k = \mathbb{C}$ and simply write $\mathbb{P}^n$ for $\mathbb{P}^n_{\mathbb{C}}$ throughout these notes.

Example 2.2. In the "projective plane" $\mathbb{P}^2$, the symbols $(2 : 0 : 1)$ and $(4 : 0 : 2)$ both refer to the same point.

It is useful to think of projective space as having its own geometric structure, rather than just as a quotient of a higher dimensional space. In particular, a **geometry** is often defined as a set along with a group of transformations. A **projective transformation** is a map $f : \mathbb{P}^n \to \mathbb{P}^n$ of the form

$$f(x_0 : x_1 : \cdots : x_n) = (y_0 : y_1 : \cdots : y_n)$$

where for each i,

$$y_i = a_{i0}x_0 + a_{i1}x_1 + \cdots a_{in}x_n$$

for some fixed constants $a_{ij} \in \mathbb{C}$ such that the $(n+1) \times (n+1)$ matrix (a_{ij}) is invertible.

Notice that projective transformations are well defined on $\mathbb{P}^n$ because scaling all the x_i variables by a constant c has the effect of scaling the y variables by c as well. This is due to the fact that the defining equations are **homogeneous**: every monomial on both sides of the equation has a fixed degree d (in this case $d = 1$).

2.1 Affine Patches and Projective Varieties

There is another way of thinking of projective space: as ordinary Euclidean space with extra smaller spaces placed out at infinity. For instance, in $\mathbb{P}^1$, any point $(x : y)$ with $y \neq 0$ can be rescaled to the form $(t : 1)$. All such points can be identified with the element $t \in \mathbb{C}$, and then there is only one more point in $\mathbb{P}^1$, namely $(1 : 0)$. We can think of $(1 : 0)$ as a point "at infinity" that closes up the *affine line* $\mathbb{C}^1$ into the "circle" $\mathbb{P}^1$. Thought of as a real surface, the complex $\mathbb{P}^1$ is actually a sphere.

Similarly, we can instead parameterize the points $(1 : t)$ by $t \in \mathbb{C}^1$ and have $(0 : 1)$ be the extra point. The subsets given by $\{(1 : t)\}$ and $\{(t : 1)\}$ are both called **affine patches** of $\mathbb{P}^1$, and form a cover of $\mathbb{P}^1$, from which we can inherit a natural topology on $\mathbb{P}^1$ from the Euclidean topology on each $\mathbb{C}^1$. In fact, the two affine patches form an open cover in this topology, so $\mathbb{P}^1$ is compact.

As another example, the projective plane $\mathbb{P}^2$ can be written as the disjoint union

$$\{(x : y : 1)\} \sqcup \{(x : 1 : 0)\} \sqcup \{1 : 0 : 0\} = \mathbb{C}^2 \sqcup \mathbb{C}^1 \sqcup \mathbb{C}^0,$$

which we can think of as a certain closure of the affine patch $\{(x : y : 1)\}$. The other affine patches are $\{(x : 1 : y)\}$ and $\{(1 : x : y)\}$ in this case.

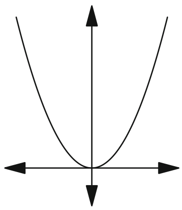
We can naturally generalize this as follows.

Definition 2.3. The **standard affine patches** of $\mathbb{P}^n$ are the sets

$$\{(t_0 : t_1 : \cdots : t_{i-1} : 1 : t_{i+1} : \cdots : t_n)\} \cong \mathbb{C}^n$$

for $i = 0, \ldots, n$.

An **affine variety** is usually defined as the set of solutions to a set of polynomials in k^n for some field k. For instance, the graph of $y = x^2$ is an affine variety in $\mathbb{R}^2$, since it is the set of all points (x, y) for which $f(x, y) = y - x^2$ is zero.



In three-dimensional space, we might consider the plane defined by the *zero locus* of $f(x, y, z) = x + y + z$, that is, the set of solutions to $f(x, y, z) = 0$. Another example is the line $x = y = z$ defined by the common zero locus of $f(x, y, z) = x - y$ and $g(x, y, z) = x - z$.

Recall that a polynomial is **homogeneous** if all of its terms have the same total degree. For instance, $x^2 + 3yz$ is homogeneous because both terms have degree 2, but $x^2 - y + 1$ is not homogeneous.

Definition 2.4. A **projective variety** is the common zero locus in $\mathbb{P}^n$ of a finite set of homogeneous polynomials $f_1(x_0, \dots, x_n), \dots, f_r(x_0, \dots, x_n)$ in $\mathbb{P}^n$. We call this variety $V(f_1, \dots, f_r)$. In other words,

$$V(f_1, \dots, f_r) = \{(a_0 : \dots : a_n) \mid f_i(a_0 : \dots : a_n) = 0 \text{ for all } i\}.$$

Remark 2.5. Note that we need the homogeneous condition in order for projective varieties to be well defined. For instance, if $f(x, y) = y - x^2$ then $f(2, 4) = 0$ and $f(4, 8) \neq 0$, but $(2 : 4) = (4 : 8)$ in $\mathbb{P}^1$. So the value of a nonhomogeneous polynomial on a point in projective space is not, in general, well defined.

The intersection of a projective variety with the ith affine patch is the *affine* variety formed by setting $x_i = 1$ in all of the defining equations. For instance, the projective variety in $\mathbb{P}^2$ defined by $f(x : y : z) = yz - x^2$ restricts to the affine variety defined by $f(x, y) = y - x^2$ in the affine patch $z = 1$.

We can also reverse this process. The **homogenization** of a polynomial $f(x_0, \dots, x_{n-1})$ in n variables using another variable x_n is the unique homogeneous polynomial $g(x_0 : \dots : x_{n-1} : x_n)$ with $\deg(g) = \deg(f)$ for which

$$g(x_0 : \dots : x_{n-1} : 1) = f(x_0, \dots, x_{n-1}).$$

For instance, the homogenization of $y - x^2$ is $yz - x^2$. If we homogenize the equations of an affine variety, we get a projective variety which we call its **projective closure**.

Example 2.6. The projective closure of the parabola defined by $y - x^2 - 1 = 0$ is the projective variety in $\mathbb{P}^3$ defined by the equation $yz - x^2 - z^2 = 0$. If we intersect this with the $y = 1$ affine patch, we obtain the affine variety $z - x^2 - z^2 = 0$ in the x, z variables. This is the circle $x^2 + (z - \frac{1}{2})^2 = \frac{1}{4}$, and so parabolas and circles are essentially the same object in projective space, cut different ways into affine patches.

As explained in more detail in Problem 2.3 below, there is only one type of (nondegenerate) conic in projective space.

Remark 2.7. The above example implies that if we draw a parabola on a large, flat plane and stand at its apex, looking out to the horizon we will see the two branches of the parabola meeting at a point on the horizon, closing up the curve into an ellipse.[4]

2.2 *Points, Lines, and m-Planes in Projective Space*

Just as the points of $\mathbb{P}^n$ are the images of lines in $\mathbb{C}^{n+1}$, a **line** in projective space can be defined as the image of a **plane** in k^{n+1}, and so on. We can define these in terms of homogeneous coordinates as follows.

[4] Unfortunately, we could not find any photographs of parabolic train tracks.

Definition 2.8. An $(n-1)$**-plane** or **hyperplane** in $\mathbb{P}^n$ is the set of solutions $(x_0 : \cdots : x_n)$ to a homogeneous linear equation

$$a_0x_0 + a_1x_1 + \cdots + a_nx_n = 0.$$

A k**-plane** is an intersection of $n-k$ hyperplanes, say $a_{i0}x_0 + a_{i1}x_1 + \cdots + a_{in}x_n = 0$ for $i = 1, \ldots, n-k$, such that the matrix of coefficients (a_{ij}) is full rank.

Example 2.9. In the projective plane $\mathbb{P}^n$, the line l_1 given by $2x + 3y + z = 0$ restricts to the line $2x + 3y + 1 = 0$ in the affine patch $z = 1$. Notice that the line l_2 given by $2x + 3y + 2z = 0$ restricts to $2x + 3y + 2 = 0$ in this affine patch, and is parallel to the restriction of l_1 in this patch. However, the projective closures of these affine lines intersect at the point $(3 : -2 : 0)$, on the $z = 0$ line at infinity.

In fact, any two distinct lines meet in a point in the projective plane. In general, intersection problems are much easier in projective space. See Problem 2.3 below to apply this to our problems in Schubert calculus.

2.3 *Problems*

2.1. **Transformations of** $\mathbb{P}^1$: Show that a projective transformation on $\mathbb{P}^1$ is uniquely determined by where it sends $0 = (0 : 1)$, $1 = (1 : 1)$, and $\infty = (1 : 0)$.

2.2. **Choice of** $n+2$ **points stabilizes** $\mathbb{P}^n$: Construct a set S of $n+2$ distinct points in $\mathbb{P}^n$ for which any projective transformation is uniquely determined by where it sends each point of S. What are necessary and sufficient conditions for a set of $n+2$ distinct points in $\mathbb{P}^n$ to have this property?

2.3. **All conics in** $\mathbb{P}^2$ **are the same**: Show that, for any quadratic homogeneous polynomial $f(x, y, z)$ there is a projective transformation that sends it to one of x^2, $x^2 + y^2$, or $x^2 + y^2 + z^2$. Conclude that any two "nondegenerate" conics are the same up to a projective transformation.

    (Hint: Any quadratic form can be written as $\mathbf{x}A\mathbf{x}^T$ where $\mathbf{x} = (x, y, z)$ is the row vector of variables and $\mathbf{x}^T$ is its transpose, and A is a symmetric matrix, with $A = A^T$. It can be shown that a symmetric matrix A can be diagonalized, i.e., written as BDB^T for some diagonal matrix D. Use the matrix B as a projective transformation to write the quadratic form as a sum of squares.)

2.4 **Schubert Calculus in Projective Space**: The question of how many points are contained in two distinct lines in $\mathbb{C}^2$ can be "projectivized" as follows: if we ask instead how many points are contained in two distinct lines in $\mathbb{P}^2$, then the answer is always 1 since parallel lines now intersect, a much nicer answer!
    Write out projective versions of Questions 1.1 and 1.2. What do they translate to in terms of intersections of subspaces of one higher dimensional affine space?

3 Theme: The Grassmannian

Not only does taking the projective closure of our problems in $\mathbb{P}^n$ make things easier, it is also useful to think of the intersection problems as involving subspaces of $\mathbb{C}^{n+1}$ rather than k-planes in $\mathbb{P}^n$. The definition of the Grassmannian below is analogous to our first definition of projective space.

Definition 3.1. The **Grassmannian** $\mathrm{Gr}(n, k)$ is the set of all k-dimensional subspaces of $\mathbb{C}^n$.

As in projective spaces, we call the elements of $\mathrm{Gr}(n, k)$ the "points" of $\mathrm{Gr}(n, k)$, even though they are defined as entire subspaces of $\mathbb{C}^n$. We will see soon that $\mathrm{Gr}(n, k)$ has the structure of a projective variety, making this notation useful.

Every point of the Grassmannian can be described as the span of some k independent row vectors of length n, which we can arrange in a $k \times n$ matrix. For instance, the matrix

$$\begin{bmatrix} 0 & -1 & -3 & -1 & 6 & -4 & 5 \\ 0 & 1 & 3 & 2 & -7 & 6 & -5 \\ 0 & 0 & 0 & 2 & -2 & 4 & -2 \end{bmatrix}$$

represents a point in $\mathrm{Gr}(7, 3)$. Notice that we can perform elementary row operations on the matrix without changing the point of the Grassmannian it represents. We will use the convention that the pivots will be in order from left to right and bottom to top.

Exercise 3.2. Show that the matrix above has reduced row echelon form:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & * & * & 0 \\ 0 & 1 & * & 0 & * & * & 0 \end{bmatrix},$$

where the $*$ entries are certain complex numbers.

We can summarize our findings as follows.

Fact 3.3. Each point of $\mathrm{Gr}(n, k)$ is the row span of a unique full-rank $k \times n$ matrix in reduced row echelon form.

The subset of the Grassmannian whose points have a particular reduced row echelon form constitutes a **Schubert cell.** Notice that $\mathrm{Gr}(n, k)$ is a disjoint union of Schubert cells.

3.1 Projective Variety Structure

The Grassmannian can be viewed as a projective variety by embedding $\mathrm{Gr}(n, k)$ in $\mathbb{P}^{\binom{n}{k}-1}$ via the *Plücker embedding*. To do so, choose an ordering on the k-element

subsets S of $\{1, 2, \ldots, n\}$ and use this ordering to label the homogeneous coordinates x_S of $\mathbb{P}^{\binom{n}{k}-1}$. Now, given a point in the Grassmannian represented by a matrix M, let x_S be the determinant of the $k \times k$ submatrix determined by the columns in the subset S. This determines a point in projective space since row operations can only change the determinants up to a constant factor, and the coordinates cannot all be zero since the matrix has rank k.

For example, in $\mathrm{Gr}(4, 2)$, the matrix

$$\begin{bmatrix} 0 & 0 & 1 & 2 \\ 1 & -3 & 0 & 3 \end{bmatrix}$$

has Plücker coordinates given by the determinants of all the 2×2 submatrices formed by choosing two of the columns above. We write x_{ij} for the determinant formed columns i and j, so for instance, $x_{24} = \det \begin{pmatrix} 0 & 2 \\ -3 & 3 \end{pmatrix} = 6$. If we order the coordinates $(x_{12} : x_{13} : x_{14} : x_{23} : x_{24} : x_{34})$ then the image of the above point under the Plücker embedding is $(0 : -1 : -2 : 3 : 6 : 3)$.

One can show that the image is a projective variety in $\mathbb{P}^{\binom{n}{k}-1}$, cut out by homogeneous quadratic relations in the variables x_S known as the *Plücker relations*. See [17], pg. 408 for details.

3.2 *Schubert Cells and Schubert Varieties*

To enumerate the Schubert cells in the Grassmannian, we assign to the matrices of the form

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & * & * & 0 \\ 0 & 1 & * & 0 & * & * & 0 \end{bmatrix}$$

a **partition**, that is, a nonincreasing sequence of nonnegative integers $\lambda = (\lambda_1, \ldots, \lambda_k)$, as follows. Cut out the $k \times k$ staircase from the upper left corner of the matrix, and let λ_i be the distance from the edge of the staircase to the 1 in row i. In the example shown, we get the partition $\lambda = (4, 2, 1)$. Notice that we always have $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$.

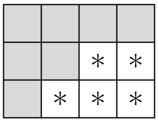
$$\begin{array}{ccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & * & * & 0 \\ 0 & 1 & * & 0 & * & * & 0 \end{array}$$

Definition 3.4. The **size** of a partition λ, denoted $|\lambda|$, is $\sum_i \lambda_i$, and its **length**, denoted $l(\lambda)$, is the number of nonzero parts. The entries λ_i are called its **parts**.

Remark 3.5. With this notation, Schubert cells in $\mathrm{Gr}(n, k)$ are in bijection with the partitions λ for which $l(\lambda) \le k$ and $\lambda_1 \le n - k$.

Definition 3.6. The **Young diagram** of a partition λ is the left-aligned partial grid of boxes in which the ith row from the top has λ_i boxes.

For example, the Young diagram of the partition $(4, 2, 1)$ that came up in the previous example is shown as the shaded boxes in the diagram below. By identifying the partition with its Young diagram, we can alternatively define λ as the complement in a $k \times (n - k)$ rectangle of the diagram μ defined by the right-aligned shift of the $*$ entries in the matrix:



Since the $k \times (n - k)$ rectangle is the bounding shape of our allowable partitions, we will call it the **ambient rectangle**.

Definition 3.7. For a partition λ contained in the ambient rectangle, the **Schubert cell** Ω_λ° is the set of points of $\mathrm{Gr}(n, k)$ whose row echelon matrix has corresponding partition λ. Explicitly,

$$\Omega_\lambda^\circ = \{V \in \mathrm{Gr}(n, k) \mid \dim(V \cap \langle e_1, \ldots, e_r \rangle) = i \text{ for } n - k + i - \lambda_i \le r \le n - k + i - \lambda_{i+1}\}.$$

Here e_{n-i+1} is the ith standard unit vector $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$ with the 1 in the ith position, so $e_1 = (0, 0, \ldots, 1)$, $e_2 = (0, 0, \ldots, 1, 0)$, and so on. The notation $\langle e_1, \ldots, e_r \rangle$ denotes the span of the vectors $e_1, \ldots, e_r$.

Remark 3.8. The numbers $n - k + i - \lambda_i$ are the positions of the 1's in the matrix counted from the right.

Since each $*$ can be any complex number, we have $\Omega_\lambda^\circ = \mathbb{C}^{k(n-k)-|\lambda|}$ as a set, and so

$$\dim(\Omega_\lambda^\circ) = k(n - k) - |\lambda|.$$

In particular, the dimension of the Grassmannian is $k(n - k)$.

We are now in a position to define **Schubert varieties** as closed subvarieties of the Grassmannian.

Definition 3.9. The **standard Schubert variety** corresponding to a partition λ, denoted Ω_λ, is the set

$$\Omega_\lambda = \{V \in \mathrm{Gr}(n, k) \mid \dim(V \cap \langle e_1, \ldots, e_{n-k+i-\lambda_i} \rangle) \ge i\}.$$

Remark 3.10. In the topology on the Grassmannian, as inherited from projective space via the Plücker embedding, the Schubert variety Ω_λ is the closure $\overline{\Omega_\lambda{}^\circ}$ of the corresponding Schubert cell. We will explore more of the topology of the Grassmannian in Sect. 4.

Note that we have $\dim(\Omega_\lambda) = \dim(\Omega_\lambda^\circ) = k(n-k) - |\lambda|$ as well.

Example 3.11. Consider the Schubert variety $\Omega_{\square\square}$ in $\mathbb{P}^5 = \mathrm{Gr}(6,1)$. The ambient rectangle is a 1×5 row of squares. There is one condition defining the points $V \in \Omega_{\square\square}$, namely $\dim(V \cap \langle e_1, e_2, e_3, e_4\rangle) \geq 1$, where V is a one-dimensional subspace of $\mathbb{C}^6$. This means that V is contained in $\langle e_1, \ldots, e_4\rangle$, and so, expressed in homogeneous coordinates, its first two entries (in positions e_5 and e_6) are 0.

Thus, each point of $\Omega_{\square\square}$ can be written in one of the following forms:

$$\begin{gathered}(0:0:1:*:*:*)\\(0:0:0:1:*:*)\\(0:0:0:0:1:*)\\(0:0:0:0:0:1)\end{gathered}$$

It follows that $\Omega_{\square\square}$ can be written as a disjoint union of Schubert cells as follows:

$$\Omega_{\square\square} = \Omega^\circ_{\square\square} \sqcup \Omega^\circ_{\square\square\square} \sqcup \Omega^\circ_{\square\square\square\square} \sqcup \Omega^\circ_{\square\square\square\square\square}.$$

In fact, every Schubert variety is a disjoint union of Schubert cells. See the problems at the end of this section for details.

We may generalize this construction to other bases than the standard basis $e_1, \ldots, e_n$, or more rigorously, using any *complete flag*. A **complete flag** is a chain of subspaces

$$F_\bullet : 0 = F_0 \subset F_1 \subset \cdots \subset F_n = \mathbb{C}^n$$

where each F_i has dimension i. Then we define

$$\Omega_\lambda(F_\bullet) = \{V \in \mathrm{Gr}(n,k) \mid \dim(V \cap F_{n-k+i-\lambda_i}) \geq i\}$$

and similarly for Ω_λ°.

Example 3.12. The Schubert variety $\Omega_\square(F_\bullet) \subset \mathrm{Gr}(4,2)$ consists of the two-dimensional subspaces V of $\mathbb{C}^4$ for which $\dim(V \cap F_2) \geq 1$. Under the quotient map $\mathbb{C}^4 \to \mathbb{P}^3$, this is equivalent to space of all lines in $\mathbb{P}^3$ that intersect a given line in at least a point, which is precisely the variety we need for Question 1.1.

3.3 A Note on Flags

Why are chains of subspaces called *flags*? Roughly speaking, a flag on a flagpole consists of:

- A point (the top of the pole),
- A line passing through that point (the pole),
- A plane passing through that line (the plane containing the flag), and
- Space to put it in.

Mathematically, this is the data of a *complete flag* in three dimensions. However, higher-dimensional beings would require more complicated flags. So in general, it is natural to define a complete flag in n-dimensional space $\mathbb{C}^n$ to be a chain of vector spaces F_i of each dimension from 0 to n, each containing the previous, with $\dim(F_i) = i$ for all i. A **partial flag** is a chain of subspaces in which only some of the possible dimensions are included.

3.4 Problems

3.1. **Projective space is a Grassmannian**: Show that every projective space $\mathbb{P}^m$ is a Grassmannian. What are n and k?

3.2. **Schubert cells in $\mathbb{P}^m$**: What are the Schubert cells in $\mathbb{P}^m$? Express your answer in homogeneous coordinates.

3.3. **Schubert varieties in $\mathbb{P}^m$**: What are the Schubert varieties in $\mathbb{P}^m$, thought of as a Grassmannian? Why are they the closures of the Schubert cells in the topology on $\mathbb{P}^m$?

3.4. **Schubert varieties versus Schubert cells**: Show that every Schubert variety is a disjoint union of Schubert cells. Describe which Schubert cells are contained in Ω_λ in terms of partitions.

3.5. **Extreme cases**: Describe $\Omega_\emptyset$ and Ω_B where B is the entire ambient rectangle. What are their dimensions?

3.6. **Intersecting Schubert Varieties**: Show that, by choosing four different flags $F_\bullet^{(1)}, F_\bullet^{(2)}, F_\bullet^{(3)}, F_\bullet^{(4)}$, Question 1.1 becomes equivalent to finding the intersection of the Schubert varieties

$$\Omega_\square(F_\bullet^{(1)}) \cap \Omega_\square(F_\bullet^{(2)}) \cap \Omega_\square(F_\bullet^{(3)}) \cap \Omega_\square(F_\bullet^{(4)}).$$

3.7. **A Variety of Varieties**: Translate the simple intersection problems of lines passing through two points, points contained in two lines, and so on into problems about intersections of Schubert varieties, as we did for Question 1.1 in Problem 3.4 above. What does Question 1.2 become?

3.8. **More complicated flag conditions**: In $\mathbb{P}^4$, let 2-planes A and B intersect in a point X, and let P and Q be distinct points different from X. Let S be the set

of all 2-planes C that contain both P and Q and intersect A and B each in a line. Express S as an intersection of Schubert varieties in $\mathrm{Gr}(5, 3)$, in each of the following cases:

(a) When P is contained in A and Q is contained in B;
(b) When neither P nor Q lie on A or B.

4 Variation 1: Intersections of Schubert Varieties in the Grassmannian

In the previous section, we saw how to express certain linear intersection problems as intersections of Schubert varieties in a Grassmannian. We now will build up the machinery needed to obtain a combinatorial rule for computing these intersections, known as the **Littlewood–Richardson rule**.

Both the geometric and combinatorial aspects of the Littlewood–Richardson rule are fairly complicated to prove, and we refer the reader to [22] for complete proofs. In this exposition, we will focus more on the applications and intuition behind the rule.

The Littlewood–Richardson rule is particularly nice in the case of zero-dimensional intersections. In particular, given a list of generic flags $F_\bullet^{(i)}$ in $\mathbb{C}^n$ for $i = 1, \ldots, r$, let $\lambda^{(1)}, \ldots, \lambda^{(r)}$ be partitions with

$$\sum |\lambda^i| = k(n-k).$$

Then the intersection

$$\bigcap \Omega_{\lambda^i}(F_\bullet^{(i)})$$

is zero-dimensional, consisting of exactly $c^B_{\lambda^{(1)},\ldots,\lambda^{(r)}}$ points of $\mathrm{Gr}(n, k)$, where B is the ambient rectangle and $c^B_{\lambda^{(1)},\ldots,\lambda^{(r)}}$ is a certain **Littlewood–Richardson coefficient**, defined in Sect. 4.6.

When we refer to a "generic" choice of flags, we mean that we are choosing from an open dense subset of the *flag variety*. This will be made more precise in Sect. 5.

In general, the Littlewood–Richardson rule computes products of Schubert classes in the cohomology ring of the Grassmannian, described in Sect. 4.4 below, which corresponds with (not necessarily zero-dimensional) intersections of Schubert varieties. To gain intuition for these intersections, we follow [22] and first simplify even further, to the case of two flags that intersect *transversely*.

4.1 Opposite and Transverse Flags, Genericity

Two subspaces of $\mathbb{C}^n$ are said to be *transverse* if their intersection has the "expected dimension". For instance, two two-dimensional subspaces of $\mathbb{C}^3$ are expected to have a one-dimensional intersection; only rarely is their intersection two-dimensional (when the two planes coincide). More rigorously:

Definition 4.1. Two subspaces V and W of $\mathbb{C}^n$ are **transverse** if

$$\dim(V \cap W) = \max(0, \dim(V) + \dim(W) - n).$$

Equivalently, if $\operatorname{codim}(V)$ is defined to be $n - \dim(V)$, then

$$\operatorname{codim}(V \cap W) = \min(n, \operatorname{codim}(V) + \operatorname{codim}(W)).$$

Exercise 4.2. Verify that the two definitions above are equivalent.

We say two flags $F_\bullet^{(1)}$ and $F_\bullet^{(2)}$ are **transverse** if every pair of subspaces $F_i^{(1)}$ and $F_j^{(2)}$ are transverse. In fact, a weaker condition suffices:

Lemma 4.3. *Two complete flags $F_\bullet, E_\bullet \subset \mathbb{C}^n$ are transverse if and only if $F_{n-i} \cap E_i = \{0\}$ for all i.*

Proof Sketch. The forward direction is clear. For the reverse implication, we can take the quotient of both flags by the one-dimensional subspace E_1 and induct on n. $\square$

Define the **standard flag** $F_\bullet$ to be the flag in which $F_i = \langle e_1, \ldots, e_i \rangle$, and similarly define the **opposite flag** $E_\bullet$ by $E_i = \langle e_n, \ldots, e_{n-i+1} \rangle$. It is easy to check that these flags $F_\bullet$ and $E_\bullet$ are transverse. Furthermore, we shall see that every pair of transverse flags can be mapped to this pair, as follows. Consider the action of $\mathrm{GL}_n(\mathbb{C})$ on $\mathbb{C}^n$ by standard matrix multiplication, and note that this gives rise to an action on flags and subspaces, and subsequently on Schubert varieties as well.

Lemma 4.4. *For any pair of transverse flags $F'_\bullet$ and $E'_\bullet$, there is an element $g \in \mathrm{GL}_n$ such that $gF'_\bullet = F_\bullet$ and $gE'_\bullet = E_\bullet$, where $F_\bullet$ and $E_\bullet$ are the standard and opposite flags.*

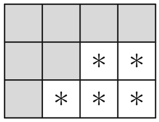
The proof of this lemma is left as an exercise to the reader (see the Problems section below). The important corollary is that to understand the intersection of the Schubert varieties $\Omega_\lambda(F'_\bullet)$ and $\Omega_\mu(E'_\bullet)$, it suffices to compute the intersection $\Omega_\lambda(F_\bullet) \cap \Omega_\lambda(E_\bullet)$ and multiply the results by the appropriate matrix g.

So, when we consider the intersection of two Schubert varieties with respect to transverse flags, it suffices to consider the standard and opposite flags $F_\bullet$ and $E_\bullet$. We use this principle in the **duality theorem** below, which tells us exactly when the intersection of $\Omega_\lambda(F_\bullet)$ and $\Omega_\mu(E_\bullet)$ is nonempty.

4.2 Duality Theorem

Definition 4.5. Two partitions $\lambda = (\lambda_1, \ldots, \lambda_k)$ and $\mu = (\mu_1, \ldots, \mu_k)$ are **complementary** in the $k \times (n-k)$ ambient rectangle if and only if $\lambda_i + \mu_{k+1-i} = n - k$ for all i. In this case we write $\mu^c = \lambda$.

In other words, if we rotate the Young diagram of μ and place it in the lower right corner of the ambient rectangle, its complement is λ. Below, we see that $\mu = (3, 2)$ is the complement of $\lambda = (4, 2, 1)$ in $\mathrm{Gr}(7, 3)$.



Theorem 4.6 (Duality Theorem). *Let $F_\bullet$ and $E_\bullet$ be transverse flags in $\mathbb{C}^n$, and let λ and μ be partitions with $|\lambda| + |\mu| = k(n-k)$. In $\mathrm{Gr}(n, k)$, the intersection $\Omega_\lambda(F_\bullet) \cap \Omega_\mu(E_\bullet)$ has 1 element if μ and λ are complementary partitions, and is empty otherwise. Furthermore, if μ and λ are any partitions with $\mu_{k+1-i} + \lambda_i > n - k$ for some i then $\Omega_\lambda(F_\bullet) \cap \Omega_\mu(E_\bullet) = \emptyset$.*

We can use a reversed form of row reduction to express the Schubert varieties with respect to the opposite flag, and then the Schubert cells for the complementary partitions will have their 1's in the same positions, as in the example below. Their unique intersection will be precisely this matrix of 1's with 0's elsewhere.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & * & * & 0 \\ 0 & 1 & * & 0 & * & * & 0 \end{bmatrix} \qquad \begin{bmatrix} * & 0 & * & 0 & * & * & 1 \\ * & 0 & * & 1 & 0 & 0 & 0 \\ * & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We now give a more rigorous proof below, which follows that in [22] but with a few notational differences.

Proof. We prove the second claim first: if for some i we have $\mu_{k+1-i} + \lambda_i > n - k$ then $\Omega_\lambda(F_\bullet) \cap \Omega_\mu(E_\bullet)$ is empty. Assume for contradiction that there is a subspace V in the intersection. We know $\dim(V) = k$, and also

$$\dim(V \cap \langle e_1, e_2, \ldots, e_{n-k+i-\lambda_i} \rangle) \geq i, \tag{1}$$

$$\dim(V \cap \langle e_n, e_{n-1}, \ldots, e_{n+1-(n-k+(k+1-i)-\mu_{k+1-i})} \rangle) \geq k + 1 - i.$$

Simplifying the last subscript above, and reversing the order of the generators, we get

$$\dim(V \cap \langle e_{i+\mu_{k+1-i}}, \ldots, e_{n-1}, e_n \rangle) \geq k + 1 - i. \tag{2}$$

Notice that $i + \mu_{k+1-i} > n - k + i - \lambda_i$ by the condition $\mu_{k+1-i} + \lambda_i > n - k$, so the two subspaces we are intersecting with V in Eqs. (1) and (2) are disjoint. It

follows that V has dimension at least $k+1-i+i=k+1$, a contradiction. Thus, $\Omega_\lambda(F_\bullet)\cap\Omega_\mu(E_\bullet)$ is empty in this case.

Thus, if $|\lambda|+|\mu|=k(n-k)$ and λ and μ are not complementary, then the intersection is empty as well, since the inequality $\mu_{k+1-i}+\lambda_i>n-k$ must hold for some i.

Finally, suppose λ and μ are complementary. Then Eqs. (1) and (2) still hold, but now $n-k+i-\lambda_i=i+\mu_{n+1-i}$ for all i. Thus $\dim(V\cap\langle e_{i+\mu_{n+1-i}}\rangle)=1$ for all $i=1,\ldots,k$, and since V is k-dimensional it must equal the span of these basis elements, namely

$$V=\langle e_{1+\mu_n},e_{2+\mu_{n-1}},\ldots e_{k+\mu_{n+1-k}}\rangle.$$

This is the unique solution. □

Example 4.7. We now can give a rather high-powered proof that there is a unique line passing through any two distinct points in $\mathbb{P}^n$. As before, we work in one higher dimensional affine space and consider 2-planes in $\mathbb{C}^{n+1}$. Working in $\mathrm{Gr}(n+1,2)$, the two distinct points become two distinct one-dimensional subspaces F_1 and E_1 of $\mathbb{C}^{n+1}$, and the Schubert condition that demands the two-dimensional subspace V contains them is

$$\dim(V\cap F_1)\geq 1,\qquad \dim(V\cap E_1)\geq 1.$$

These are the Schubert conditions for a single-part partition $\lambda=(\lambda_1)$ where $(n+1)-2+1-\lambda_1=1$. Thus $\lambda_1=n-1$, and we are intersecting the Schubert varieties

$$\Omega_{(n-1)}(F_\bullet)\cap\Omega_{(n-1)}(E_\bullet)$$

where $F_\bullet$ and $E_\bullet$ are any two transverse flags extending F_1 and E_1. Notice that $(n-1)$ and $(n-1)$ are complementary partitions in the $2\times(n-1)$ ambient rectangle (see Fig. 2), so by the Duality Theorem there is a unique point of $\mathrm{Gr}(n+1,2)$ in the intersection. The conclusion follows.

4.3 Cell Complex Structure

In order to prove the more general zero-dimensional Littlewood–Richardson rule and compute the Littlewood–Richardson coefficients, we need to develop more heavy machinery. In particular, we need to understand the Grassmannian as a geometric object and compute its **cohomology**, an associated ring in which multiplication of

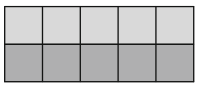


Fig. 2 Two complimentary partitions of size $n-1$ filling the $n-1\times 2$ rectangle

certain generators will correspond to intersection of Schubert varieties. (See [27] for more details on all of the material in this section.)

The term *Schubert cell* comes from the notion of a **cell complex** (also known as a CW complex) in algebraic topology. An n**-cell** is a topological space homeomorphic to the open ball $|v| < 1$ in $\mathbb{R}^n$, and its associated n**-disk** is its closure $|v| \le 1$ in $\mathbb{R}^n$.

To construct a cell complex, one starts with a set of points called the 0**-skeleton** X^0, and then attaches 1-disks D via continuous boundary maps from the boundary ∂D (which consists of two points) to X^0. The result is a 1**-skeleton** X^1.

This can then be extended to a 2-skeleton by attaching 2-disks D via maps from the boundary ∂D (which is a circle) to X^1. In general, the n-skeleton X^n is formed by attaching a set of n-disks to X^{n-1} along their boundaries.

More precisely, to form X^n from X^{n-1}, we start with a collection of n-disks D^n_α and continuous attaching maps $\varphi_\alpha : \partial D^n_\alpha \to X^{n-1}$. Then

$$X^n = \frac{X^{n-1} \sqcup \bigsqcup_\alpha D^n_\alpha}{\sim}$$

where $\sim$ is the identification $x \sim \varphi_\alpha(x)$ for $x \in \partial D^n_\alpha$. The cell complex is $X = \bigcup_n X^n$, which may be simply $X = X^n$ if the process stops at stage n. By the construction, the points of X^0 along with the open i-cells associated with the i-disks in X^i for each i are disjoint and cover the cell complex X. The topology is given by the rule that $A \subset X$ is open if and only if $A \cap X^n$ is open in X^n for all n, where the topology on X^n is given by the usual Euclidean topology on $\mathbb{R}^n$.

Example 4.8. The real projective plane $\mathbb{P}^2_{\mathbb{R}}$ has a cell complex structure in which $X^0 = \{(0:0:1)\}$ is a single point, $X^1 = X^0 \sqcup \{(0:1:*)\}$ is topologically a circle formed by attaching a 1-cell to the point at both ends, and then X^2 is formed by attaching a 2-cell $\mathbb{R}^2$ to the circle such that the boundary wraps around the 1-cell twice. This is because the points of the form $(1 : xt : yt)$ as $t \to \infty$ and as $t \to -\infty$ both approach the same point in X^1, so the boundary map must be a 2-to-1 mapping.

Example 4.9. The *complex* projective plane $\mathbb{P}^2_{\mathbb{C}}$ has a simpler cell complex structure, consisting of starting with a single point $X^0 = \{(0:0:1)\}$, and then attaching a 2-cell (a copy of $\mathbb{C} = \mathbb{R}^2$) like a balloon to form X^2. A copy of $\mathbb{C}^2 = \mathbb{R}^4$ is then attached to form X^4.

The Schubert cells give a cell complex structure on the Grassmannian. For a complete proof of this, see [53], section 3.2. We sketch the construction below.

Define the 0-skeleton X^0 to be the zero-dimensional Schubert variety $\Omega_{((n-k)^k)}$. Define X^2 to be X^0 along with the 2-cell (since we are working over $\mathbb{C}$ and not $\mathbb{R}$) given by $\Omega^\circ_{((n-k)^{k-1},n-k-1)}$, and the attaching map given by the closure in $\mathrm{Gr}(n,k)$. Note that the partition in this step is formed by removing a single corner square from the ambient rectangle.

Then, X^4 is formed by attaching the two four-cells given by removing two outer corner squares in both possible ways, giving either $\Omega^\circ_{((n-k)^{k-2},n-k-1,n-k-1)}$ or

$\Omega^\circ_{((n-k)^{k-1},n-k-2)}$. We can continue in this manner with each partition size to define the entire cell structure, $X^0 \subset X^2 \subset \cdots \subset X^{2k(n-k)}$.

Example 4.10. We have

$$\mathrm{Gr}(4,2) = \Omega^\circ_{\ydiagram{2,2}} \sqcup \Omega^\circ_{\ydiagram{2,1}} \sqcup \Omega^\circ_{\ydiagram{2}} \sqcup \Omega^\circ_{\ydiagram{1,1}} \sqcup \Omega^\circ_{\ydiagram{1}} \sqcup \Omega^\circ_{\emptyset},$$

forming a cell complex structure in which $X^0 = \Omega^\circ_{\ydiagram{2,2}}$, X^2 is formed by attaching $\Omega^\circ_{\ydiagram{2,1}}$, X^4 is formed by attaching $\Omega^\circ_{\ydiagram{2}} \sqcup \Omega^\circ_{\ydiagram{1,1}}$, X^6 is formed by attaching $\Omega^\circ_{\ydiagram{1}}$, and X^8 is formed by attaching $\Omega^\circ_{\emptyset}$.

4.4 Cellular Homology and Cohomology

For a CW complex $X = X^0 \subset \cdots \subset X^n$, define

$$C_k = \mathbb{Z}^{\#k\text{-cells}},$$

the free abelian group generated by the k-cells $B^{(k)}_\alpha = (D^{(k)}_\alpha)^\circ$.

Define the **cellular boundary map** $d_{k+1} : C_{k+1} \to C_k$ by

$$d_{k+1}(B^{(k+1)}_\alpha) = \sum_\beta \deg_{\alpha\beta} \cdot B^{(k)}_\beta,$$

where $\deg_{\alpha\beta}$ is the *degree* of the composite map

$$\partial \overline{B^{(k+1)}_\alpha} \to X^k \to \overline{B^{(k)}_\beta}.$$

The first map above is the cellular attaching map from the boundary of the closure of the ball $B^{(k+1)}_\alpha$ to the k-skeleton, and the second map is the quotient map formed by collapsing $X^k \setminus B^{(k)}_\beta$ to a point. The composite is a map from a k-sphere to another k-sphere, which has a **degree**, whose precise definition we omit here and refer the reader to [27], section 2.2, p. 134. As one example, the 2-to-1 attaching map described in Example 4.8 for $\mathbb{P}^2_{\mathbb{R}}$ has degree 2.

It is known that the cellular boundary maps make the groups C_k into a **chain complex**: a sequence of maps

$$0 \to C_n \xrightarrow{d_n} C_{n-1} \xrightarrow{d_{n-1}} C_{n-2} \to \cdots \to C_1 \xrightarrow{d_1} C_0 \to 0$$

for which $d_i \circ d_{i+1} = 0$ for all i. This latter condition implies that the image of the map d_{i+1} is contained in the kernel of d_i for all i, and so we can consider the quotient groups

$$H_i(X) = \ker(d_i)/\mathrm{im}(d_{i+1})$$

for all i. These quotients are abelian groups called the **cellular homology groups** of the space X.

Example 4.11. Recall that $\mathbb{P}^2_{\mathbb{C}}$ consists of a point, a 2-cell, and a 4-cell. So, its cellular chain complex is:

$$\cdots \to 0 \to 0 \to 0 \to \mathbb{Z} \to 0 \to \mathbb{Z} \to 0 \to \mathbb{Z} \to 0$$

and the homology groups are $H_0 = H_2 = H_4 = \mathbb{Z}$, $H_1 = H_3 = 0$.

On the other hand, in $\mathbb{P}^2_{\mathbb{R}}$, the chain complex looks like:

$$0 \to \mathbb{Z} \to \mathbb{Z} \to \mathbb{Z} \to 0$$

where the first map $\mathbb{Z} \to \mathbb{Z}$ is multiplication by 2 and the second is the 0 map, due to the degrees of the attaching maps. It follows that $H_2 = 0$, $H_1 = \mathbb{Z}/2\,\mathbb{Z}$, and $H_0 = \mathbb{Z}$.

We can now define the **cellular cohomology** by dualizing the chain complex above. In particular, define

$$C^k = \mathrm{Hom}(C_k, \mathbb{Z}) = \{\text{group homomorphisms } f : C_k \to \mathbb{Z}\}$$

for each k, and define the **coboundary maps** $d_k^* : C^{k-1} \to C^k$ by

$$d_k^* f(c) = f(d_k(c))$$

for any $f \in C^k$ and $c \in C_k$. Then the coboundary maps form a **cochain complex**, and we can define the cohomology groups to be the abelian groups

$$H^i(X) = \ker(d_{i+1}^*)/\mathrm{im}(d_i^*)$$

for all i.

Example 4.12. The cellular cochain complex for $\mathbb{P}^2_{\mathbb{C}}$ is

$$0 \to \mathbb{Z} \to 0 \to \mathbb{Z} \to 0 \to \mathbb{Z} \to 0 \to 0 \to 0 \to \cdots$$

and so the cohomology groups are $H^0 = H^2 = H^4 = \mathbb{Z}$, $H^1 = H^3 = 0$.

Finally, the direct sum of the cohomology groups

$$H^*(X) = \bigoplus H^i(X)$$

has a ring structure given by the **cup product** ([27], p. 249), which is the dual of the "cap product" ([27], p. 239) on homology and roughly corresponds to taking intersection of cohomology classes in this setting.

In particular, there is an equivalent definition of cohomology on the Grassmannian known as the *Chow ring*, in which cohomology classes in $H^*(X)$ are equivalence classes of algebraic subvarieties under **birational equivalence**. (See [21], Sections 1.1 and 19.1.) In other words, deformations under rational families are still equivalent: in $\mathbb{P}^2$, for instance, the family of algebraic subvarieties of the form $xy - tz^2 = 0$ as $t \in \mathbb{C}$ varies are all in one equivalence class, even as $t \to 0$ and the hyperbola degenerates into two lines.

The main fact we will be using under this interpretation is the following, which we state without proof. (See [22], Section 9.4 for more details.)

Theorem 4.13. *The cohomology ring $H^*(\mathrm{Gr}(n,k))$ has a $\mathbb{Z}$-basis given by the classes*

$$\sigma_\lambda := [\Omega_\lambda(F_\bullet)] \in H^{2|\lambda|}(\mathrm{Gr}(n,k))$$

for λ a partition fitting inside the ambient rectangle. The cohomology $H^*(\mathrm{Gr}(n,k))$ is a graded ring, so $\sigma_\lambda \cdot \sigma_\mu \in H^{2|\lambda|+2|\mu|}(\mathrm{Gr}(n,k))$, and we have

$$\sigma_\lambda \cdot \sigma_\mu = [\Omega_\lambda(F_\bullet) \cap \Omega_\mu(E_\bullet)]$$

where $F_\bullet$ and $E_\bullet$ are the standard and opposite flags.

Note that σ_λ is independent of the choice of flag $F_\bullet$, since any two Schubert varieties of the same partition shape are rationally equivalent via a change of basis.

We can now restate the intersection problems in terms of multiplying Schubert classes. In particular, if $\lambda^{(1)}, \ldots, \lambda^{(r)}$ are partitions with $\sum_i |\lambda^{(i)}| = k(n-k)$, then

$$\sigma_{\lambda^{(1)}} \cdots \sigma_{\lambda^{(r)}} \in H^{k(n-k)}(\mathrm{Gr}(n,k))$$

and there is only one generator of the top cohomology group, namely σ_B where B is the ambient rectangle. This is the cohomology class of the single point $\Omega_B(F_\bullet)$ for some flag $F_\bullet$. Thus the intersection of the Schubert varieties $\Omega_{\lambda^{(1)}}(F_\bullet^{(1)}) \cap \cdots \cap \Omega_{\lambda^{(r)}}(F_\bullet^{(r)})$ is rationally equivalent to a finite union of points, the number of which is the coefficient $c^B_{\lambda^{(1)},\ldots,\lambda^{(r)}}$ in the expansion

$$\sigma_{\lambda^{(1)}} \cdots \sigma_{\lambda^{(r)}} = c^B_{\lambda^{(1)},\ldots,\lambda^{(r)}} \sigma_B.$$

For a sufficiently general choice of flags, the $c^B_{\lambda^{(1)},\ldots,\lambda^{(r)}}$ points in the intersection are distinct with no multiplicity.

In general, we wish to understand the coefficients that we get upon multiplying Schubert classes and expressing the product back in the basis $\{\sigma_\lambda\}$ of Schubert classes.

Example 4.14. In Problem 3.4, we saw that Question 1.1 can be rephrased as computing the size of the intersection

$$\Omega_\square(F_\bullet^{(1)}) \cap \Omega_\square(F_\bullet^{(2)}) \cap \Omega_\square(F_\bullet^{(3)}) \cap \Omega_\square(F_\bullet^{(4)})$$

for a given generic choice of flags $F_\bullet^{(1)}, \ldots, F_\bullet^{(4)}$. By the above analysis, we can further reduce this problem to computing the coefficient c for which

$$\sigma_\square \cdot \sigma_\square \cdot \sigma_\square \cdot \sigma_\square = c \cdot \sigma_{\boxplus}$$

in $H^*(\mathrm{Gr}(4, 2))$.

4.5 Connection with Symmetric Functions

We can model the cohomology ring $H^*(\mathrm{Gr}(n, k))$ algebraically as a quotient of the ring of **symmetric functions**. We only cover the essentials of symmetric function theory for our purposes here, and refer the reader to Chapter 7 of [51], or the books [38] or [48] for more details, or to [22] for the connection between $H^*(\mathrm{Gr}(n, k))$ and the ring of symmetric functions.

Definition 4.15. The ring of *symmetric functions* $\Lambda_{\mathbb{C}}(x_1, x_2, \ldots)$ is the ring of bounded-degree formal power series $f \in \mathbb{C}[[x_1, x_2, \ldots]]$ which are symmetric under permuting the variables, that is,

$$f(x_1, x_2, \ldots) = f(x_{\pi(1)}, x_{\pi(1)}, \ldots)$$

for any permutation $\pi : \mathbb{Z}_+ \to \mathbb{Z}_+$ and $\deg(f) < \infty$.

For instance, $x_1^2 + x_2^2 + x_3^2 + \cdots$ is a symmetric function of degree 2.

The most important symmetric functions for Schubert calculus are the *Schur functions*. They can be defined in many equivalent ways, from being characters of irreducible representations of S_n to an expression as a ratio of determinants. We use the combinatorial definition here and start by introducing some common terminology involving Young tableaux and partitions.

Definition 4.16. A **skew shape** is the difference ν/λ formed by cutting out the Young diagram of a partition λ from the strictly larger partition ν. A skew shape is a **horizontal strip** if no column contains more than one box.

Definition 4.17. A **semistandard Young tableau (SSYT)** of shape ν/λ is a way of filling the boxes of the Young diagram of ν/λ with positive integers so that the numbers are weakly increasing across rows and strictly increasing down columns. An SSYT has **content** μ if there are μ_i boxes labeled i for each i. The **reading word** of the tableau is the word formed by concatenating the rows from bottom to top.

The following is a semistandard Young tableau of shape ν/λ and content μ where $\nu = (6, 5, 3)$, $\lambda = (3, 2)$, and $\mu = (4, 2, 2, 1)$. Its reading word is 134223111.

			1	1	1
		2	2	3	
1	3	4			

Definition 4.18. Let λ be a partition. Given a semistandard Young tableau T of shape λ, define $x^T = x_1^{m_1} x_2^{m_2} \cdots$ where m_i is the number of i's in T. The **Schur function** for a partition λ is the symmetric function defined by

$$s_\lambda = \sum_T x^T$$

where the sum ranges over all SSYT's T of shape λ.

Example 4.19. For $\lambda = (2, 1)$, the tableaux

$$\begin{array}{|c|c|}\hline 1 & 1 \\ \hline 2 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 2 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 1 & 1 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array} \quad \cdots$$

are a few of the infinitely many SSYT's of shape λ. Thus we have

$$s_\lambda = x_1^2 x_2 + x_1 x_2^2 + x_1^2 x_3 + 2x_1 x_2 x_3 + \cdots .$$

It is well known that the Schur functions s_λ are symmetric and form a vector space basis of $\Lambda(x_1, x_2, \ldots)$ as λ ranges over all partitions. The key fact that we will need is that they allow us to understand the cohomology ring $H^*(\mathrm{Gr}(n, k))$, as follows.

Theorem 4.20. *There is a ring isomorphism*

$$H^*(\mathrm{Gr}(n, k)) \cong \Lambda(x_1, x_2, \ldots)/(s_\lambda | \lambda \not\subset B)$$

where B is the ambient rectangle and $(s_\lambda | \lambda \not\subset B)$ is the ideal generated by the Schur functions. The isomorphism sends the Schubert class σ_λ to the Schur function s_λ.

This is a pivotal theorem in the study of the Grassmannian, since it allows us to compute in the cohomology ring simply by working with symmetric polynomials. In particular, multiplying Schur functions corresponds to multiplying cohomology classes, which in turn gives us information about intersections of Schubert varieties.

As an approach to prove Theorem 4.20, note that sending σ_λ to s_λ is an isomorphism of the underlying vector spaces, since on the right-hand side we have quotiented by the Schur functions whose partition does not fit inside the ambient rectangle. So, it remains to show that this isomorphism respects the multiplications in these rings, taking cup product to polynomial multiplication.

An important first step is the **Pieri Rule**. For Schur functions, this tells us how to multiply a one-row shape by any other partition:

$$s_{(r)} \cdot s_\lambda = \sum_{\nu/\lambda \text{ horz. strip of size } r} s_\nu .$$

We wish to show that the same relation holds for the σ_λ's, that is, that

$$\sigma_{(r)} \cdot \sigma_\lambda = \sum_{\nu/\lambda \text{ horz. strip of size } r} \sigma_\nu,$$

where the sum on the right is restricted to partitions ν fitting inside the ambient rectangle. Note that we do not need this restriction for general Schur functions, but in the cohomology ring we are considering the quotient by partitions not fitting inside the ambient rectangle, so the two expansions above are not exactly the same.

Note that, by the Duality Theorem, we can multiply both sides of the above relation by σ_{μ^c} to extract the coefficient of σ_μ on the right-hand side. So, the Pieri Rule is equivalent to the following restatement:

Theorem 4.21 (Pieri Rule)**.** *Let λ and μ be partitions with $|\lambda| + |\mu| = k(n-k) - r$. Then if $F_\bullet$, $E_\bullet$, and $H_\bullet$ are three sufficiently general flags then the intersection*

$$\Omega_\lambda(F_\bullet) \cap \Omega_\mu(E_\bullet) \cap \Omega_{(r)}(H_\bullet)$$

has 1 *element if μ^c/λ is a horizontal strip, and it is empty otherwise.*

Sketch of Proof. We can set $F_\bullet$ and $E_\bullet$ to be the standard and opposite flags and $H_\bullet$ a generic flag distinct from $F_\bullet$ or $E_\bullet$. We can then perform a direct analysis similar to that in the Duality Theorem. See [22] for full details. □

Algebraically, the Pieri rule suffices to show the ring isomorphism, because the Schur functions $s_{(r)}$ and corresponding Schubert classes $\sigma_{(r)}$ form an algebraic set of generators for their respective rings. Therefore, to intersect Schubert classes we simply have to understand how to multiply Schur functions.

4.6 The Littlewood–Richardson Rule

The combinatorial rule for multiplying Schur functions, or Schubert classes, is called the **Littlewood–Richardson Rule**. To state it, we need to introduce a few new notions.

Definition 4.22. A word $w_1 w_2 \cdots w_n$ (where each $w_i \in \{1, 2, 3, \ldots\}$) is **Yamanouchi** (or **lattice** or **ballot**) if every suffix $w_k w_{k+1} \cdots w_n$ contains at least as many letters equal to i as $i + 1$ for all i.

For instance, the word 231211 is Yamanouchi, because the suffixes 1, 11, 211, 1211, 31211, and 231211 each contain at least as many 1's as 2's, and at least as many 2's as 3's.

Definition 4.23. A **Littlewood–Richardson tableau** is a semistandard Young tableau whose reading word is Yamanouchi.

Exercise 4.24. The example tableau in Fig. 3 is **not** Littlewood–Richardson. Why? Can you find a tableau of that shape that is?

Fig. 3 An example of a skew Littlewood–Richardson tableau

			1	1	1
		2	2	3	
1	3	4			

Definition 4.25. A sequence of skew tableaux $T_1, T_2, \ldots$ form a **chain** if their shapes do not overlap and

$$T_1 \cup T_2 \cup \cdots \cup T_i$$

is a partition shape for all i.

We can now state the general Littlewood–Richardson rule. We will refer the reader to [22] for a proof, as the combinatorics is quite involved.

Theorem 4.26. *We have*

$$s_{\lambda^{(1)}} \cdot \cdots \cdot s_{\lambda^{(m)}} = \sum_{\nu} c^{\nu}_{\lambda^{(1)},\ldots,\lambda^{(m)}} s_{\nu}$$

where $c^{\nu}_{\lambda^{(1)},\ldots,\lambda^{(m)}}$ is the number of chains of Littlewood–Richardson tableaux of contents $\lambda^{(i)}$ with total shape ν.

It is worth noting that in many texts, the following corollary is the primary focus, since the above theorem can be easily derived from the $m = 2$ case stated below.

Corollary 4.27. *We have*

$$s_{\lambda} s_{\mu} = \sum_{\nu} c^{\nu}_{\lambda\mu} s_{\nu}$$

where $c^{\nu}_{\lambda\mu}$ is the number of Littlewood–Richardson tableaux of skew shape ν/λ and content μ.

Proof. By Theorem 4.26, $c^{\nu}_{\lambda\mu}$ is the number of chains of two Littlewood–Richardson tableaux of content λ and μ with total shape ν. The first tableau of content λ is a straight shape tableau, so by the Yamanouchi reading word condition and the semistandard condition, the top row can only contain 1's. Continuing this reasoning inductively, it has only i's in its ith row for each i. Therefore, the first tableau in the chain is the unique tableau of shape λ and content λ.

Thus, the second tableau is a Littlewood–Richardson tableau of shape ν/λ and content μ, and the result follows. □

As a consequence of Theorems 4.26 and 4.20, in $H^*(\mathrm{Gr}(n, k))$ we have

$$\sigma_{\lambda^{(1)}} \cdot \cdots \cdot \sigma_{\lambda^{(m)}} = \sum_{\nu} c^{\nu}_{\lambda^{(1)},\ldots,\lambda^{(m)}} \sigma_{\nu}$$

where now the sum on the right is restricted to partitions ν fitting inside the ambient rectangle. Note that by the combinatorics of the Littlewood–Richardson rule, the coefficients on the right are nonzero only if $|\nu| = \sum |\lambda^{(i)}|$, and so in the case of a zero-dimensional intersection of Schubert varieties, the only possible ν on the right-hand side is the ambient rectangle B itself. Moreover, $\Omega_B(F_\bullet)$ is a single point of $\mathrm{Gr}(n,k)$ for any flag $F_\bullet$. The zero-dimensional Littlewood–Richardson rule follows as a corollary.

Theorem 4.28 (Zero-Dimensional Littlewood–Richardson Rule)**.** *Let B be the $k \times (n-k)$ ambient rectangle, and let $\lambda^{(1)}, \dots, \lambda^{(m)}$ be partitions fitting inside B such that $|B| = \sum_i |\lambda^{(i)}|$. Also let $F_\bullet^{(1)}, \dots, F_\bullet^{(m)}$ be any m generic flags. Then*

$$c^B_{\lambda^{(1)},\dots,\lambda^{(m)}} := |\Omega_{\lambda^{(1)}}(F_\bullet^{(1)}) \cap \cdots \cap \Omega_{\lambda^{(m)}}(F_\bullet^{(m)})|$$

is equal to the number of chains of Littlewood–Richardson tableaux of contents $\lambda^{(1)}, \dots, \lambda^{(m)}$ with total shape equal to B.

Example 4.29. Suppose $k = 3$ and $n - k = 4$. Let $\lambda^{(1)} = (2, 1)$, $\lambda^{(2)} = (2, 1)$, $\lambda^{(3)} = (3, 1)$, and $\lambda^{(4)} = 2$. Then there are five different chains of Littlewood–Richardson tableaux of contents $\lambda^{(1)}, \dots, \lambda^{(4)}$ that fill the $k \times (n-k)$ ambient rectangle, as shown in Fig. 4. Thus $c^B_{\lambda^{(1)},\dots,\lambda^{(4)}} = 5$.

Example 4.30. We can now solve Question 1.1. In Example 4.14, we showed that it suffices to compute the coefficient c in the expansion

$$\sigma_\square \cdot \sigma_\square \cdot \sigma_\square \cdot \sigma_\square = c \cdot \sigma_{\boxplus}$$

in $H^*(\mathrm{Gr}(4,2))$. This is the Littlewood–Richardson coefficient $c^{(2,2)}_{\square,\square,\square,\square}$. This is the number of ways to fill a 2×2 ambient rectangle with a chain of Littlewood–Richardson tableaux having one box each.

Since such a tableau can only have a single 1 as its entry, we will label the entries with subscripts indicating the step in the chain to distinguish them. We have two possibilities, as shown in Fig. 5. Therefore the coefficient is 2, and so there are 2 lines passing through four generic lines in $\mathbb{P}^4$.

In Example 4.30, we are in the special case in which each Littlewood–Richardson tableau in the chain has only one box, and so the only choice we have is the ordering

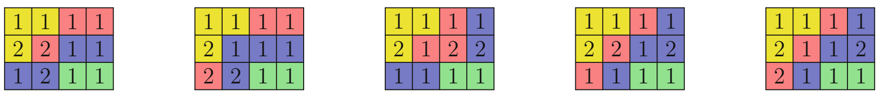


Fig. 4 The five chains of Littlewood–Richardson tableaux of contents $\lambda^{(1)} = (2, 1)$, $\lambda^{(2)} = (2, 1)$, $\lambda^{(3)} = (3, 1)$, and $\lambda^{(4)} = 2$ filling an ambient 3×4 rectangle

$$\begin{array}{|c|c|}\hline 1_1 & 1_2 \\ \hline 1_3 & 1_4 \\ \hline \end{array} \qquad\qquad \begin{array}{|c|c|}\hline 1_1 & 1_3 \\ \hline 1_2 & 1_4 \\ \hline \end{array}$$

Fig. 5 The two tableaux chains used to enumerate the Littlewood–Richardson coefficient that answers Question 1.1

of the boxes in a way that forms a chain. We can, therefore, simply represent such a tableau by its indices instead, and the two tableaux of Fig. 5 become

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 4 \\ \hline \end{array} \quad \text{and} \quad \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & 4 \\ \hline \end{array}.$$

The two tableaux above are characterized by the property that the entries $1, 2, 3, 4$ are used exactly once and the rows and columns are strictly increasing. Such a tableau is called a *standard Young tableaux*.

Definition 4.31. A **standard Young tableau** of shape λ with $|\lambda| = n$ is an SSYT of shape λ in which the numbers $1, 2, \ldots, n$ are each used exactly once.

There is a well-known explicit formula, known as the **Hook length formula**, for the number of standard Young tableaux of a given shape, due to Frame, Robinson, and Thrall [18]. To state it we need the following definition.

Definition 4.32. For a square s in a Young diagram, define the **hook length**

$$\mathrm{hook}(s) = \mathrm{arm}(s) + \mathrm{leg}(s) + 1$$

where $\mathrm{arm}(s)$ is the number of squares strictly to the right of s in its row and $\mathrm{leg}(s)$ is the number of squares strictly below s in its column.

Theorem 4.33 *(Hook length formula.) The number of standard Young tableaux of shape λ is*

$$\frac{|\lambda|!}{\prod_{s\in\lambda} \mathrm{hook}(s)}.$$

For example, if $\lambda = (2, 2)$ then we have $\frac{4!}{3\cdot 2\cdot 2\cdot 1} = 2$ standard Young tableaux of shape λ, which matches our answer in Example 4.30.

4.7 Problems

4.1. **Prove Lemma 4.4:** For any transverse flags $F'_\bullet$ and $E'_\bullet$, there is some $g \in \mathrm{GL}_n$ such that $gF'_\bullet = F_\bullet$ and $gE'_\bullet = E_\bullet$, where $F_\bullet$ and $E_\bullet$ are the standard and opposite flags.

4.2. **It's all Littlewood–Richardson:** Verify that the Duality Theorem and the Pieri Rule are both special cases of the Littlewood–Richardson rule.

4.3. **An empty intersection:** Show that

$$\Omega_{(1,1)}(F_\bullet) \cap \Omega_{(2)}(E_\bullet) = \emptyset$$

in $\mathrm{Gr}(4, 2)$ for transverse flags $F_\bullet$ and $E_\bullet$. What does this mean geometrically?

4.4. **A nonempty intersection:** Show that

$$\Omega_{(1,1)}(F_\bullet) \cap \Omega_{(2)}(E_\bullet)$$

is nonempty in $\mathrm{Gr}(5, 2)$. (Hint: intersecting it with a certain third Schubert variety will be nonempty by the Littlewood–Richardson rule.) What does this mean geometrically?

4.5. **Problem 3.4 revisited:** In $\mathbb{P}^4$, suppose the 2-planes A and B intersect in a point X, and P and Q are distinct points different from X. Show that there is exactly one plane C that contains both P and Q and intersect A and B each in a line as an intersection of Schubert varieties in $\mathrm{Gr}(5, 3)$, in each of the following cases:

   (a) When P is contained in A and Q is contained in B;
   (b) When neither P nor Q lie on A or B.

4.6. **That's a lot of k-planes:** Solve Question 1.2 for a generic choice of flags as follows.

   (a) Verify that the problem boils down to computing the coefficient of $s_{((n-k)^k)}$ in the product of Schur functions $s_{(1)}^{k(n-k)}$.
   (b) Use the Hook Length Formula to finish the computation.

5 Variation 2: The Flag Variety

For the content in this section, we refer to [39], unless otherwise noted below.

The (complete) **flag variety** (in dimension n) is the set of all complete flags in $\mathbb{C}^n$, with a Schubert cell decomposition similar to that of the Grassmannian. In particular, given a flag

$$V_\bullet : V_0 \subset V_1 \subset \cdots V_n = \mathbb{C}^n,$$

we can choose n vectors $v_1, \ldots, v_n$ such that the span of $v_1, \ldots, v_i$ is V_i for each i, and list the vectors v_i as row vectors of an $n \times n$ matrix. We can then perform certain row reduction operations to form a different ordered basis $v_1', \ldots, v_n'$ that also span the subspaces of the flag, but whose matrix entries consist of a permutation matrix of 1's, all 0's to the left and below each 1, and arbitrary complex numbers in all other entries.

For instance, say we start with the flag in three dimensions generated by the vectors $(0, 2, 3)$, $(1, 1, 4)$, and $(1, 2, -3)$. The corresponding matrix is

$$\begin{pmatrix} 0 & 2 & 3 \\ 1 & 1 & 4 \\ 1 & 2 & -3 \end{pmatrix}.$$

We start by finding the leftmost nonzero element in the first row and scale that row so that this element is 1. Then subtract multiples of this row from the rows below it so that all the entries below that 1 are 0. Continue the process on all further rows:

$$\begin{pmatrix} 0 & 2 & 3 \\ 1 & 1 & 4 \\ 1 & 2 & -3 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & 1.5 \\ 1 & 0 & 2.5 \\ 1 & 0 & -6 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & 1.5 \\ 1 & 0 & 2.5 \\ 0 & 0 & 1 \end{pmatrix}$$

It is easy to see that this process does not change the flag formed by the initial row spans and that any two matrices in canonical form define different flags. So, the flag variety is a cell complex consisting of $n!$ **Schubert cells** indexed by permutations. For instance, one such open set in the five-dimensional flag variety is the open set given by all matrices of the form

$$\begin{pmatrix} 0 & 1 & * & * & * \\ 1 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & * & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

We call this cell X°_{45132} because $4, 5, 1, 3, 2$ are the positions of the 1's from the right-hand side of the matrix in order from top to bottom. More rigorously, we define a Schubert cell as follows.

Definition 5.1 Let $w \in S_n$ be a permutation of $\{1, \ldots, n\}$. Then the **Schubert cell** of w is defined by

$$X^\circ_w = \{V_\bullet \in \mathrm{Fl}_n : \dim(V_p \cap F_q) = \#\{i \le p : w(i) \le q\} \text{ for all } p, q\}$$

where $F_\bullet$ is the standard flag generated by the unit vectors e_{n+1-i}. In the matrix form above, the columns are ordered from right to left as before.

Note that, as in the case of the Grassmannian, we can choose a different flag $F_\bullet$ with respect to which we define our Schubert cell decomposition, and we define $X^\circ_w(F_\bullet)$ accordingly.

The dimension of a Schubert cell X_w is the number of $*$'s in its matrix, that is, the number of entries above and right of the pivot 1 in its row and column. The maximum number of $*$'s occurs when the permutation is $w_0 = n(n-1)\cdots 321$, in which case the dimension of the open set X_{w_0} is $n(n-1)/2$ (or $n(n-1)$ over $\mathbb{R}$). In general,

it is not hard to see that the number of $*$'s in the set X_w is the **inversion number** $\mathrm{inv}(w)$. This is defined to be the number of pairs of entries $(w(i), w(j))$ of w which are out of order, that is, $i < j$ and $w(i) > w(j)$. Thus we have

$$\dim(X_w^\circ) = \mathrm{inv}(w).$$

Example 5.2 The permutation $w = 45132$ has seven inversions. (Can you find them all?) We also see that $\dim(X_w^\circ) = 7$, since there are seven $*$ entries in the matrix.

Another useful way to think of $\mathrm{inv}(w)$ is in terms of its **length**.

Definition 5.3 Define $s_1, \ldots, s_{n-1} \in S_n$ to be the *adjacent transpositions* in the symmetric group, that is, s_i is the permutation interchanging i and $i+1$. Then the **length** of w, written $\ell(w)$, is the smallest number k for which there exists a decomposition

$$w = s_{i_1} \cdots s_{i_k}.$$

Lemma 5.4 *We have* $\ell(w) = \mathrm{inv}(w)$ *for any* $w \in S_n$.

We will leave the proof of this lemma as an exercise to the reader in the Problems Section.

5.1 *Schubert Varieties and the Bruhat Order*

By using the Plücker embeddings $\mathrm{Gr}(n, k) \hookrightarrow \mathbb{P}^{\binom{n}{k}-1}$ for each k, we can embed Fl_n into the larger projective space $\mathbb{P}^{2^n-1}$ whose entries correspond to the Plücker coordinates of each of the initial $k \times n$ submatrices of a given element of the flag variety. This makes Fl_n a projective subvariety of $\mathbb{P}^{2^n-1}$ (see [22] for more details), which in turn gives rise to a topology on Fl_n, known as the Zariski topology. Now, consider the closures of the sets X_w° in this topology.

Definition 5.5 The **Schubert variety** corresponding to a permutation $w \in S_n$ is

$$X_w = \overline{X_w^\circ}.$$

As in the Grassmannian, these Schubert varieties turn out to be disjoint unions of Schubert cells. The partial ordering in which $X_w = \sqcup_{v \le w} X_v^\circ$ is called the **Bruhat order**, a well-known partial order on permutations. We will briefly review it here, but we refer to [10] for an excellent introduction to Bruhat order.

Definition 5.6 The **Bruhat order** $\le$ on S_n is defined by $v \le w$ if and only if, for every representation of w as a product of $l(w)$ transpositions s_i, one can remove $l(w) - l(v)$ of the transpositions to obtain a representation of v as a subword in the same relative order.

Example 5.7 The permutation $w = 45132$ can be written as $s_2s_3s_2s_1s_4s_3s_2$. This contains $s_3s_2s_3 = 14325$ as a (non-consecutive) subword, and so $14325 \leq 45132$.

5.2 *Intersections and Duality*

Now suppose we wish to answer incidence questions about our flags: which flags satisfy certain linear constraints? As in the case of the Grassmannian, this boils down to understanding how the Schubert varieties X_w intersect.

We start with the Duality Theorem for Fl_n. Following [22], it will be convenient to define dual Schubert varieties as follows.

Definition 5.8 Let $E_\bullet$ be the standard and opposite flags, and for shorthand we let $X_w = X_w(F_\bullet)$ and

$$Y_w = X_{w_0 \cdot w}(E_\bullet)$$

where $w_0 = n(n-1)\cdots 1$ is the longest word. The set Y_w is often called a **dual Schubert variety**.

Notice that

$$\dim(Y_w) = \mathrm{inv}(w_0 \cdot w) = n(n-1)/2 - \mathrm{inv}(w)$$

since if $w' = w_0 \cdot w$ then $w'(i) = n + 1 - w(i)$ for all i.

Theorem 5.9 (Duality Theorem, V2.)**.** *If $l(w) = l(v)$, we have $X_w \cap Y_v = \emptyset$ if $w \neq v$ and $|X_w \cap Y_v| = 1$ if $w = v$. Furthermore, if $l(w) < l(v)$ then $X_w \cap Y_v = \emptyset$.*

The proof works similarly to the Duality Theorem in the Grassmannian. In particular, with respect to the standard basis, the dual Schubert variety Y_w is formed by the same permutation matrix of 1's as in X_w, but with the 0 entries below and to the *right* of the 1's (and $*$ entries elsewhere). For instance, we have

$$X_{45132} = \begin{pmatrix} 0 & 1 & * & * & * \\ 1 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & * & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \qquad Y_{45132} = \begin{pmatrix} * & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & * & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

and their intersection contains only the permutation matrix determined by $w = 45132$.

5.3 Schubert Polynomials and the Cohomology Ring

In order to continue our variation on the theme, it would be natural at this point to look for a Pieri rule or a Littlewood–Richardson rule. But just as the cohomology ring of the Grassmannian and the Schur functions made those rules more natural, we now turn to **Schubert polynomials** and the cohomology ring $H^*(\mathrm{Fl}_n)$ over $\mathbb{Z}$.

This ring has a natural interpretation as a quotient of a polynomial ring. In particular, letting σ_w be the cohomology class of Y_w, we have $\sigma_w \in H^{2i}(\mathrm{Fl}_n)$ where $i = \mathrm{inv}(w)$. For the transpositions s_i, we have $\sigma_{s_i} \in H^2(\mathrm{Fl}_n)$. The elements $x_i = \sigma_{s_i} - \sigma_{s_{i+1}}$ for $i \leq n-1$ and $x_n = -\sigma_{s_{n-1}}$ gives a set of generators for the cohomology ring, and in fact

$$H^*(\mathrm{Fl}_n) = \mathbb{Z}[x_1, \ldots, x_n]/(e_1, \ldots, e_n) =: R_n$$

where $e_1, \ldots, e_n$ are the elementary symmetric polynomials in $x_1, \ldots, x_n$. (See [22] or [6].)

The ring R_n is known as the **coinvariant ring** and arises in many geometric and combinatorial contexts. Often defined over a field k rather than $\mathbb{Z}$, its dimension as a k-vector space (or rank as a $\mathbb{Z}$-module) is $n!$. There are many natural bases for R_n of size $n!$, such as the monomial basis given by

$$\{x_1^{a_1} \cdots x_n^{a_n} : a_i \leq n-i \text{ for all } i\}$$

(see, for instance [23]), the harmonic polynomial basis (see [5], Section 8.4) and the Schubert basis described below. There are also many famous generalizations of the coinvariant ring, such as the Garsia–Procesi modules [23] and the diagonal coinvariants (see [5], Chapter 10), which are closely tied to the study of Macdonald polynomials in symmetric function theory [38].

The Schubert polynomials form a basis of R_n whose product corresponds to the intersection of Schubert varieties. To define them, we require a divided difference operator.

Definition 5.10. For any polynomial $P(x_1, \ldots, x_n) \in \mathbb{Z}[x_1, \ldots, x_n]$, we define

$$\partial_i(P) = \frac{P - s_i(P)}{x_i - x_{i+1}}$$

where $s_i(P) = P(x_1, \ldots, x_{i-1}, x_{i+1}, x_i, x_{i+2}, \ldots, x_n)$ is the polynomial formed by switching x_i and x_{i+1} in P.

We can use these operators to recursively define the Schubert polynomials.

Definition 5.11. We define the Schubert polynomials $\mathfrak{S}_w$ for $w \in S_n$ by:

- $\mathfrak{S}_{w_0} = x_1^{n-1} x_2^{n-2} \cdots x_{n-2}^2 x_{n-1}$ where $w_0 = n(n-1)\cdots 21$ is the longest permutation,

- If $w \neq w_0$, find a minimal factoriation of the form $w = w_0 \cdot s_{i_1} \cdot \cdots \cdot s_{i_r}$ is a minimal factorization of its form, that is, a factorization for which $\ell(w_0 \cdot s_{i_1} \cdot \cdots \cdot s_{i_p}) = n - p$ for all $1 \leq p \leq r$. Then

$$\mathfrak{S}_w = \partial_{i_r} \circ \partial_{i_{r-1}} \circ \cdots \circ \partial_{i_1}(\mathfrak{S}_{w_0})$$

Remark 5.12. One can show that the operators ∂_i satisfy the two relations below.

- **Commutation Relation:** $\partial_i\partial_j = \partial_j\partial_i$ for any i, j with $|i - j| > 1$,
- **Braid Relation:** $\partial_i\partial_{i+1}\partial_i = \partial_{i+1}\partial_i\partial_{i+1}$ for any i.

Since these (along with $s_i^2 = 1$) generate all relations satisfied by the reflections s_i (see chapter 3 of [10]), the construction in Definition 5.11 is independent of the choice of minimal factorization. Note also that $\partial_i^2 = 0$, so the requirement of minimal factorizations is necessary in the definition.

The Schubert polynomials' image in R_n not only form a basis of these cohomology rings, but the polynomials themselves form a basis of all polynomials in the following sense. The Schubert polynomials $\mathfrak{S}_w$ are well defined for permutations $w \in S_\infty = \bigcup S_m$ for which $w(i) > w(i+1)$ for all $i \geq k$ for some k. For a fixed such k, these Schubert polynomials form a basis for $\mathbb{Z}[x_1, \ldots, x_k]$.

One special case of the analog of the Pieri rule for Schubert polynomials is known as **Monk's rule**.

Theorem 5.13 (Monk's rule)**.** *We have*

$$\mathfrak{S}_{s_i} \cdot \mathfrak{S}_w = \sum \mathfrak{S}_v$$

where the sum ranges over all permutations v obtained from w by:

- *Choosing a pair p, q of indices with $p \leq i < q$ for which $w(p) < w(q)$ and for any k between p and q, $w(k)$ is not between $w(p)$ and $w(q)$,*
- *Defining $v(p) = w(q)$, $v(q) = w(p)$ and for all other k, $v(k) = w(k)$.*

Equivalently, the sum is over all $v = w \cdot t$ where t is a transposition (pq) with $p \leq i < q$ for which $l(v) = l(w) + 1$.

Interestingly, there is not a known "Littlewood–Richardson rule" that generalizes Monk's rule, and this is an important open problem in Schubert calculus.

Open Problem 5.14. *Find a combinatorial interpretation analogous to the Littlewood–Richardson rule for the positive integer coefficients $c_{u,v}^w$ in the expansion*

$$\mathfrak{S}_u \cdot \mathfrak{S}_v = \sum c_{u,v}^w \mathfrak{S}_w,$$

and therefore for computing the intersection of Schubert varieties in Fl_n.

Similar open problems exist for other *partial flag varieties*, defined in the next sections.

5.4 Two Alternative Definitions

There are two other ways of defining the flag manifold that are somewhat less explicit but more generalizable. The group $\mathrm{GL}_n = \mathrm{GL}_n(\mathbb{C})$ acts on the set of flags by left multiplication on its ordered basis. Under this action, the stabilizer of the standard flag $F_\bullet$ is the subgroup B consisting of all invertible upper triangular matrices. Notice that GL_n acts transitively on flags via change-of-basis matrices, and so the stabilizer of any arbitrary flag is simply a conjugation gBg^{-1} of B. We can, therefore, define the flag variety as the set of cosets in the quotient GL_n/B, and define its variety structure accordingly.

Alternatively, we can associate to each coset gB in GL_n/B the subgroup gBg^{-1}. Since B is its own *normalizer* in G ($gBg^{-1} = B$ iff $g \in B$), the cosets in GL_n/B are in one-to-one correspondence with subgroups conjugate to B. We can, therefore, define the flag variety as the set $\mathcal{B}$ of all subgroups conjugate to B.

5.5 Generalized Flag Varieties

The notion of a "flag variety" can be extended in an algebraic way starting from the definition as GL_n/B, to quotients of other matrix groups G by certain subgroups B called **Borel subgroups**. The subgroup B of invertible upper triangular matrices is an example of a Borel subgroup of GL_n, that is, a **maximal connected solvable subgroup**. It is *connected* because it is the product of the torus $(\mathbb{C}^*)^n$ and $\binom{n}{2}$ copies of $\mathbb{C}$. We can also show that it is *solvable*, meaning that its derived series of commutators

$$\begin{aligned} B_0 &:= B, \\ B_1 &:= [B_0, B_0], \\ B_2 &:= [B_1, B_1], \\ &\vdots \end{aligned}$$

terminates. Indeed, $[B, B]$ is the set of all matrices of the form $bcb^{-1}c^{-1}$ for b and c in B. Writing $b = (d_1 + n_1)$ and $c = (d_1 + n_2)$ where d_1 and d_2 are diagonal matrices and n_1 and n_2 strictly upper triangular, it is not hard to show that $bcb^{-1}c^{-1}$ has all 1's on the diagonal. By a similar argument, one can show that the elements of B_2 have 1's on the diagonal and 0's on the off-diagonal, and B_3 has two off-diagonal rows of 0's, and so on. Thus, the derived series is eventually the trivial group.

In fact, a well-known theorem of Lie and Kolchin [31] states that *all* solvable subgroups of GL_n consist of upper triangular matrices in some basis. This implies that B is maximal as well among solvable subgroups. Therefore, B is a Borel subgroup.

The Lie–Kolchin theorem also implies that all the Borel subgroups in GL_n are of the form gBg^{-1} (and all such groups are Borel subgroups). That is, all Borel subgroups are conjugate. It turns out that this is true for any **semisimple linear**

algebraic group G, that is, a matrix group defined by polynomial equations in the matrix entries, such that G has no nontrivial smooth connected solvable normal subgroups.

Additionally, any Borel subgroup in a semisimple linear algebraic group G is its own normalizer. By an argument identical to that in the previous section, it follows that the groups G/B are independent of the choice of Borel subgroup B (up to isomorphism) and are also isomorphic to the set $\mathcal{B}$ of all Borel subgroups of G as well. Therefore, we can think of $\mathcal{B}$ as an algebraic variety by inheriting the structure from G/B for any Borel subgroup B.

Finally, we can define a generalized flag variety as follows.

Definition 5.15. The **flag variety** of a semisimple linear algebraic group G to be G/B where B is a Borel subgroup.

Some classical examples of such linear algebraic groups are the special linear group SL_n, the special orthogonal group SO_n of orthogonal $n \times n$ matrices, and the symplectic group SP_{2n} of symplectic matrices. We will explore a related quotient of the special orthogonal group SO_{2n+1} in Sect. 6.

We now define partial flag varieties, another generalization of the complete flag variety. Recall that a **partial flag** is a sequence $F_{i_1} \subset \cdots \subset F_{i_r}$ of subspaces of $\mathbb{C}^n$ with $\dim(F_{i_j}) = i_j$ for all j. Notice that a k-dimensional subspace of $\mathbb{C}^n$ can be thought of as a partial flag consisting of a single subspace F_k.

It is not hard to show that all **partial flag varieties**, the varieties of partial flags of certain degrees, can be defined as a quotient G/P for a **parabolic subgroup** P, namely a closed intermediate subgroup $B \subset P \subset G$. The Grassmannian $\mathrm{Gr}(n, k)$, then, can be thought of as the quotient of GL_n by the parabolic subgroup $S = \mathrm{Stab}(V)$ where V is any fixed k-dimensional subspace of $\mathbb{C}^n$. Similarly, we can start with a different algebraic group, say the special orthogonal group SO_{2n+1}, and quotient by parabolic subgroups to get partial flag varieties of other types.

5.6 Problems

5.1. **Reflection length equals inversion number:** Show that $l(w) = \mathrm{inv}(w)$ for any $w \in S_n$.
5.2. **Practice makes perfect:** Write out all the Schubert polynomials for permutations in S_3 and S_4.
5.3. **Braid relations:** Verify that the operators ∂_i satisfy the braid relations as stated in Remark 5.12.
5.4. **The product rule for Schubert calculus:** Prove that $\partial_i(P \cdot Q) = \partial_i(P) \cdot Q + s_i(P) \cdot \partial_i(Q)$ for any two polynomials P and Q.
5.5. **Divided difference acts on** R_n: Use the previous problem to show that the operator ∂_i maps the ideal generated by elementary symmetric polynomials to itself, and hence, the operator descends to a map on the quotient R_n.

5.6. **Schubert polynomials as a basis:** Prove that if $w \in S_\infty$ satisfies $w(i) > w(i+1)$ for all $i \geq k$ then $\mathfrak{S}_w \in \mathbb{Z}[x_1, \ldots, x_k]$. Show that they form a basis of the polynomial ring as well.

6 Variation 3: The Orthogonal Grassmannian

In the previous section, we saw that we can interpret the Grassmannian as a partial flag variety. We can generalize this construction to other matrix groups G, hence, defining Grassmannians in other Lie types. We will explore one of these Grassmannians as our final variation.

Definition 6.1. The **orthogonal Grassmannian** $\mathrm{OG}(2n+1, k)$ is the quotient SO_{2n+1}/P where P is the stabilizer of a fixed **isotropic** k-dimensional subspace V. The term *isotropic* means that V satisfies $\langle v, w\rangle = 0$ for all $v, w \in V$ with respect to a chosen symmetric bilinear form $\langle , \rangle$.

The isotropic condition, at first glance, seems very unnatural. After all, how could a nonzero subspace possibly be orthogonal to itself? Well, it is first important to note that we are working over $\mathbb{C}$, not $\mathbb{R}$, and the bilinear form is symmetric, not conjugate-symmetric. In particular, suppose we define the bilinear form to be the usual dot product

$$\langle (a_1, \ldots, a_{2n+1}), (b_1, \ldots, b_{2n+1})\rangle = a_1 b_1 + a_2 b_2 + \cdots + a_{2n+1} b_{2n+1}$$

in $\mathbb{C}^{2n+1}$. Then in $\mathbb{C}^3$, the vector $(3, 5i, 4)$ is orthogonal to itself: $3 \cdot 3 + 5i \cdot 5i + 4 \cdot 4 = 0$.

While the choice of symmetric bilinear form does not change the fundamental geometry of the orthogonal Grassmannian, one choice in particular makes things easier to work with in practice: the "reverse dot product" given by

$$\langle (a_1, \ldots, a_{2n+1}), (b_1, \ldots, b_{2n+1})\rangle = \sum_{i=1}^{2n+1} a_i b_{2n+1-i}.$$

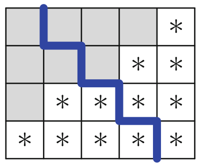
In particular, with respect to this symmetric form, the standard complete flag $F_\bullet$ is an **orthogonal flag**, with $F_i^\perp = F_{2n+1-i}$ for all i. Orthogonal flags are precisely the type of flags that are used to define Schubert varieties in the orthogonal Grassmannian.

Note that isotropic subspaces are sent to other isotropic subspaces under the action of the orthogonal group: if $\langle v, w\rangle = 0$ then $\langle Av, Aw\rangle = \langle v, w\rangle = 0$ for any $A \in \mathrm{SO}_{2n+1}$. Thus the orthogonal Grassmannian $\mathrm{OG}(2n+1, k)$, which is the quotient $\mathrm{SO}_{2n+1}/\mathrm{Stab}(V)$, can be interpreted as the variety of all k-dimensional isotropic subspaces of $\mathbb{C}^{2n+1}$.

6.1 Schubert Varieties and Row Reduction in $\mathbf{OG}(2n+1, n)$

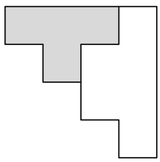
Just as in the ordinary Grassmannian, there is a Schubert cell decomposition for the orthogonal Grassmannian. The combinatorics of Schubert varieties are particularly nice in the case of $\mathrm{OG}(2n+1, n)$ in which the orthogonal subspaces are "half dimension" n. (See the introduction of [54] or the book [26] for more details.)

In $\mathrm{Gr}(2n+1, n)$, the Schubert varieties are indexed by partitions λ whose Young diagram fit inside the $n \times (n+1)$ ambient rectangle. Suppose we divide this rectangle into two staircases as shown below using the blue cut, and only consider the partitions λ that are symmetric with respect to the reflective map taking the upper staircase to the lower.



We claim that the Schubert varieties of the orthogonal Grassmannian are indexed by the **shifted partitions** formed by ignoring the lower half of these symmetric partition diagrams. We define the **ambient triangle** to be the half of the ambient rectangle above the staircase cut.

Definition 6.2. A **shifted partition** is a strictly decreasing sequence of positive integers, $\lambda = (\lambda_1 > \ldots > \lambda_k)$. We write $|\lambda| = \sum \lambda_i$. The **shifted Young diagram** of λ is the partial grid in which the ith row contains λ_i boxes and is shifted to the right i steps. Below is the shifted Young diagram of the shifted partition (3, 1), drawn inside the ambient triangle from the example above.



Definition 6.3. Let $F_\bullet$ be an orthogonal flag in $\mathbb{C}^{2n+1}$, and let λ be a shifted partition. Then the **Schubert variety** $X_\lambda(F_\bullet)$ is defined by

$$X_\lambda(F_\bullet) = \{W \in \mathrm{OG}(2n+1, n) : \dim(W \cap F_{n+1+i-\overline{\lambda}_i}) \geq i \text{ for } i = 1, \ldots, n\}$$

where $\overline{\lambda}$ is the "doubled partition" formed by reflecting the shifted partition about the staircase.

In other words, the Schubert varieties consist of the isotropic elements of the ordinary Schubert varieties, giving a natural embedding $\mathrm{OG}(2n+1, n) \to \mathrm{Gr}(2n+1, n)$ that respects the Schubert decompositions:

$$X_\lambda(F_\bullet) = \Omega_{\overline{\lambda}}(F_\bullet) \cap \mathrm{OG}(2n+1, n).$$

To get a sense of how this works, consider the example of $\lambda = (3, 1)$ and $\overline{\lambda} = (4, 3, 1)$ shown above, in the case $n = 4$. The Schubert cell $\Omega^\circ_{\overline{\lambda}}$ in $\mathrm{Gr}(9, 4)$ looks like

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & 0 & * \\ 0 & 0 & 1 & * & * & 0 & * & 0 & * \\ 1 & * & 0 & * & * & 0 & * & 0 & * \end{bmatrix}$$

Now, which of these spaces are isotropic? Suppose we label the starred entries as shown, omitting the 0 entries:

$$\begin{array}{c@{\quad}c} \begin{bmatrix} & & & & & & & 1 & l \\ & & & & & 1 & j & & k \\ & & 1 & f & g & & h & & i \\ 1 & a & & b & c & & d & & e \end{bmatrix} & \begin{matrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \end{matrix} \end{array}$$

We will show that the entries l, j, k, h, i, e are all uniquely determined by the values of the remaining variables a, b, c, d, f, g. Thus, there is one isotropic subspace in this cell for each choice of values a, b, c, d, f, g, corresponding to the "lower half" of the partition diagram we started with, namely

$$\begin{array}{|c|c|c|c|c|} \hline & & & & l \\ \hline & & & j & k \\ \hline & f & g & h & i \\ \hline a & b & c & d & e \\ \hline \end{array}$$

To see this, let the rows of the matrix be labeled **1**, **2**, **3**, **4** from top to bottom as shown, and suppose its row span is isotropic. Since row **1** and **4** are orthogonal with respect to the reverse dot product, we get the relation

$$l + a = 0,$$

which expresses $l = -a$ in terms of a.

Rows **2** and **4** are also orthogonal, which means that

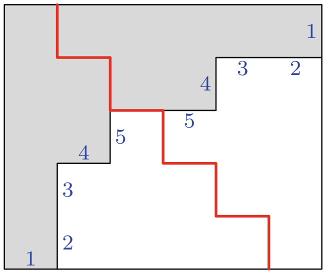
$$b + k = 0,$$

so we can similarly eliminate k. From rows **2** and **3**, we obtain $f + j = 0$, which expresses j in terms of the lower variables. We then pair row **3** with itself to see that $h + g^2 = 0$, eliminating h, and finally pairing **3** with **4** we have $i + gc + d = 0$, so i is now expressed in terms of lower variables as well.

Moreover, these are the only relations we get from the isotropic condition—any other pairings of rows give the trivial relation $0 = 0$. So in this case the Schubert variety restricted to the orthogonal Grassmannian has half the dimension of the original, generated by the possible values for a, b, c, d, f, g.

6.2 *General Elimination Argument*

Why does the elimination process work for any symmetric shape λ? Label the steps of the boundary path of λ by $1, 2, 3, \ldots$ from SW to NE in the lower left half, and label them from NE to SW in the upper right half, as shown:



Then the labels on the vertical steps in the lower left half give the column indices of the 1's in the corresponding rows of the matrix. The labels on the horizontal steps in the upper half, which match these labels by symmetry, give the column indices *from the right* of the corresponding starred columns from right to left.

$$
\begin{array}{c}
\begin{array}{cccccccccc} 2 & 3 & & 5 & & & & & & \end{array} \\
\left[\begin{array}{cccccccccc}
 & & & & & & & & & 1 \\
 & & & & & & 1 & * & * & \\
 & & & 1 & * & * & & * & * & \\
 & 1 & * & & * & * & & * & * & \\
1 & & * & & * & * & & * & * &
\end{array}\right] \\
\begin{array}{cccccccccc} & & & & & 5 & & 3 & 2 & \end{array}
\end{array}
$$

This means that the 1's in the lower left of the matrix correspond to the opposite columns of those containing letters in the upper right half. It follows that we can use the orthogonality relations to pair a 1 (which is leftmost in its row) with a column entry in a higher or equal row so as to express that entry in terms of other letters to its lower left. The 1 is in a lower or equal row in these pairings precisely for the entries whose corresponding square lies above the staircase cut. Thus, we can always express the upper right variables in terms of the lower left, as in our example above.

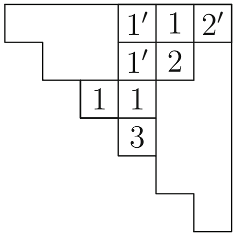


Fig. 6 The tableau above is a shifted semistandard tableau of shape λ/μ where $\lambda = (6, 4, 2, 1)$ and $\mu = (3, 2)$, and content $(5, 2, 1)$. Its reading word is $3111'21'12'$

6.3 *Shifted Tableaux and a Littlewood–Richardson Rule*

The beauty of shifted partitions is that so much of the original tableaux combinatorics that goes into ordinary Schubert calculus works almost the same way for shifted tableaux and the orthogonal Grassmannian. We define these notions rigorously below.

Definition 6.4. A **shifted semistandard Young tableau** is a filling of the boxes of a shifted skew shape with entries from the alphabet $\{1' < 1 < 2' < 2 < 3' < 3 < \cdots\}$ such that the entries are weakly increasing down columns and across rows, and such that primed entries can only repeat in columns, and unprimed only in rows.

The **reading word** of such a tableau is the word formed by concatenating the rows from bottom to top. The **content** of T is the vector $\mathrm{content}(T) = (n_1, n_2, \ldots)$, where n_i is the total number of (i)s and (i')s in T. See Fig. 6 for an example.

In this setting, there are actually two analogs of "Schur functions" that arise from these semistandard tableaux. They are known as the Schur P-functions and Schur Q-functions.

Definition 6.5. Let λ/μ be a shifted skew shape. Define $\mathrm{ShST}_Q(\lambda/\mu)$ to be the set of all shifted semistandard tableaux of shape λ/μ. Define $\mathrm{ShST}_P(\lambda/\mu)$ to be the set of those tableaux in which primes are not allowed on the staircase diagonal.

Definition 6.6. The **Schur Q-function** $Q_{\lambda/\mu}$ is defined as

$$Q_{\lambda/\mu}(x_1, x_2, \ldots) = \sum_{T \in \mathrm{ShST}_Q(\lambda/\mu)} x^{\mathrm{wt}(T)}$$

and the **Schur P-function** $P_{\lambda/\mu}$ is defined as

$$P_{\lambda/\mu}(x_1, x_2, \ldots) = \sum_{T \in \mathrm{ShST}_P(\lambda/\mu)} x^{\mathrm{wt}(T)}.$$

The Schur Q-functions, like ordinary Schur functions, are symmetric functions with unique leading terms, spanning a proper subspace of Λ. In addition, they have positive product expansions

$$Q_\mu Q_\nu = \sum 2^{\ell(\mu)+\ell(\nu)-\ell(\lambda)} f_{\mu\nu}^\lambda Q_\lambda$$

for certain positive integers $f_{\mu\nu}^\lambda$. It is easy to see that this is equivalent to the rule

$$P_\mu P_\nu = \sum f_{\mu\nu}^\lambda P_\lambda.$$

Here the coefficients $f_{\mu\nu}^\lambda$ are precisely the structure coefficients for the cohomology ring of the orthogonal Grassmannian. In particular, if we extend them to generalized coefficients by

$$P_{\mu^{(1)}} \cdot \cdots \cdot P_{\mu^{(r)}} = \sum f_{\mu^{(1)}\dots\mu^{(r)}}^\lambda P_\lambda,$$

we have the following theorem due to Pragacz [44].

Theorem 6.7. *A zero-dimensional intersection $X_{\mu^{(1)}} \cap \cdots \cap X_{\mu^{(r)}}$ has exactly $f_{\mu^{(1)}\dots\mu^{(r)}}^T$ points, where T is the ambient triangle.*

Stembridge [52] first found a Littlewood–Richardson-type rule to enumerate these coefficients. The rule is as follows.

Definition 6.8. Let T be a semistandard shifted skew tableau with the first i or i' in reading order unprimed, and with reading word $w = w_1 \cdots w_n$. Let $m_i(j)$ be the multiplicity of i among $w_{n-j+1}, \dots, w_n$ (the last j entries) for any i and for any $j \le n$. Also let $p_i(j)$ be the multiplicity of i' among $w_1, \dots, w_j$. Then T is **Littlewood–Richardson** if and only if

- Whenever $m_i(j) = m_{i+1}(j)$ we have $w_{n-j} \neq i+1, (i+1)'$, and
- Whenever $m_i(n) + p_i(j) = m_{i+1}(n) + p_i(j)$ we have $w_{j+1} \neq i, (i+1)'$.

Notice that this definition implies that $m_i(j) \ge m_{i+1}(j)$ for all i and j, which is similar to the usual Littlewood–Richardson definition for ordinary tableaux. An alternative rule that only requires reading through the word once (rather than once in each direction, as in the definition of m_i above) is given in [25].

6.4 Problems

6.1. Show that, if λ is a partition that is *not* symmetric about the staircase cut, the intersection $\Omega_\lambda^\circ(F_\bullet) \cap \mathrm{OG}(2n+1, n)$ is empty.

6.2. How many isotropic 3-planes in $\mathbb{C}^7$ intersect six given 3-planes each in at least dimension 1?

7 Conclusion and Further Variations

In this exposition, we have only explored the basics of the cohomology of the Grassmannian, the complete flag variety, and the orthogonal Grassmannian. There are many other natural directions one might explore from here.

First and foremost, we recommend that interested readers next turn to Fulton's book entitled Young Tableaux [22] for more details on the combinatorial aspects of Schubert calculus and symmetric functions, including connections with representation theory. Other books that are a natural next step from this exposition are those of Manivel [39], Kumar on Kac–Moody groups and their flag varieties [33], and Billey–Lakshmibai on smoothness and singular loci of Schubert varieties [9].

In some more specialized directions, the flag varieties and Grassmannians in other Lie types (as briefly defined in Sect. 6) have been studied extensively. The combinatorics of general Schubert polynomials for other Lie types was developed by Billey and Haiman in [8] and also by Fomin and Kirillov in type B [20]. Combinatorial methods for minuscule and cominuscule types is presented in [54].

It is also natural to investigate partial flag varieties between the Grassmannian and Fl_n. Buch, Kresch, Purbhoo, and Tamvakis established a Littlewood–Richardson rule in the special case of *two-step* flag varieties (consisting of the partial flags having just two subspaces) in [13], and the three-step case was very recently solved by Knutson and Zinn-Justin [30]. Coskun provided a potential alternative approach in terms of *Mondrian tableaux*, with a full preliminary answer for partial flag varieties in [15], and for the two-row case in [16].

Other variants of cohomology, such as *equivariant cohomology* and *K-theory*, have been extensively explored for the Grassmannian and the flag variety as well. An excellent introduction to equivariant cohomology can be found in [2, 12] is a foundational paper on the K-theory of Grassmannians. The K-theoretic analog of Schubert polynomials are called *Grothendieck polynomials*, first defined by Lascoux and Schutzenberger [36].

Another cohomological variant is *quantum cohomology*, originally arising in string theory and put on mathematical foundations in the 1990s (see [32, 46]). Fomin, Gelfand, and Postnikov [19] studied a quantum analog of Schubert polynomials and their combinatorics. Chen studied quantum cohomology on flag manifolds in [14], and the case of equivariant quantum cohomology has been more recently explored by Anderson and Chen in [3] and Bertiger, Milićević, and Taipale in [7]. In [41, 42], Pechenik and Yong prove a conjecture of Knutson and Vakil that gives a rule for equivariant K-theory of Grassmannians. The list goes on; there are many cohomology theories (in fact, infinitely many, in some sense) all of which give slightly different insight into the workings of Grassmannians and flag varieties.

It is worth noting that Young tableaux are not the only combinatorial objects that can be used to describe these cohomology theories. Knutson, Tao, and Woodward developed the theory of *puzzles* in [29], another such combinatorial object which often arises in the generalizations listed above.

On the geometric side, Vakil [56] discovered a "geometric Littlewood–Richardson Rule" that describes an explicit way to degenerate an intersection of Schubert varieties into a union of other Schubert varieties (not just at the level of cohomology). This, in some sense, more explicitly answers the intersection problems described in Sect. 1.

Another natural geometric question is the smoothness and singularities of Schubert varieties. Besides the book by Billey and Lakshmibai mentioned above [9], this has been studied for the full flag variety by Lakshmibai and Sandya [35], in which they found a pattern avoidance criterion on permutations w for which the Schubert variety X_w is smooth. Related results on smoothness in partial flag varieties and other variants have been studied by Gasharov and Reiner [24], Ryan [47], and Wolper [57]. Abe and Billey [1] summarized much of this work with a number of results on pattern avoidance in Schubert varieties.

Real Schubert calculus (involving intersection problems in real n-dimensional space $\mathbb{R}^n$) is somewhat more complicated than the complex setting, but there are still many nice results in this area. For instance, a theorem of Mukhin, Tarasov, and Varchenko in [40] states that for a choice of flags that are each maximally tangent at some real point on the rational normal curve, the intersections of the corresponding *complex* Schubert varieties have all real solutions. An excellent recent overview of this area was written by Sottile in [50].

Relatedly, one can study the *positive* real points of the Grassmannian, that is, the subset of the Grassmannian whose Plücker coordinates have all positive (or nonnegative) real values. Perhaps one of the most exciting recent developments is the connection with scattering amplitudes in quantum physics, leading to the notion of an *amplituhedron* coming from a positive Grassmannian. An accessible introduction to the main ideas can be found in [11], and for more in-depth study, the book [4] by Arkani-Hamed et al. In [43], Postnikov, Speyer, and Williams explore much of the rich combinatorial foundations of the positive Grassmannian.

Finally, there are also many geometric spaces that have some similarities with the theory of Grassmannians and flag varieties. *Hessenberg varieties* are a family of subvarieties of the full flag variety determined by stability conditions under a chosen linear transformations (see Tymoczko's thesis [55], for instance). Lee studied the combinatorics of the affine flag variety in detail in [37]. The book *k-Schur functions and affine Schubert calculus* by Lam, Lapointe, Morse, Schilling, Shimozono, and Zabrocki [34] gives an excellent overview of this area, its connections to k-Schur functions, and the unresolved conjectures on their combinatorics.

Acknowledgements The author thanks Jennifer Morse, François Bergeron, Franco Saliola, and Luc Lapointe for the invitation to teach a graduate workshop on Schubert calculus at the Center for Mathematics Research in Montreal, for which these extended notes were written. Thanks also to Sara Billey and the anonymous reviewer for their extensive feedback. Thanks to Helene Barcelo, Sean Griffin, Philippe Nadeau, Alex Woo, Jake Levinson, and Guanyu Li for further suggestions and comments. Finally, thanks to all of the participants at the graduate workshop for their comments, questions, and corrections that greatly improved this exposition.

References

1. H. Abe, S. Billey, Consequences of the Lakshmibai-Sandhya Theorem: The ubiquity of permutation patterns in Schubert calculus and related geometry (2014), arxiv:1403.4345
2. D. Anderson, Introduction to equivariant cohomology in algebraic geometry, Notes on lectures by W. Fulton at IMPAGNA summer school (2010), arXiv:1112.1421
3. D. Anderson, L. Chen, Equivariant quantum Schubert polynomials. Adv. Math. **254**, 300–330 (2014)
4. N. Arkani-Hamed, J.L. Bourjaily, F. Cachazo, A.B. Goncharov, A. Postnikov, J. Trnka, *Grassmannian Geometry of Scattering Amplitudes* (Cambridge University Press, Cambridge, 2016)
5. F. Bergeron, *Algebraic Combinatorics and Coinvariant Spaces* (CRC Press, Baco Raton, 2009)
6. I.N. Bernstein, I.M. Gelfand, S.I. Gelfand, Schubert cells and cohomology of the spaces G/P. Russ. Math. Surv. **28**(3), 1–26 (1973)
7. A. Bertiger, E. Milićević, K. Taipale, Equivariant quantum cohomology of the Grassmannian via the rim hook rule (2014), arxiv:1403.6218
8. S. Billey, M. Haiman, Schubert polynomials for the classical groups. J. Amer. Math. Soc. **8**(2) (1995)
9. S. Billey, V. Lakshmibai, *Singular Loci of Schubert Varieties* (Springer, Berlin, 2000)
10. A. Bjorner, F. Brenti, *Combinatorics of Coxeter Groups* (Springer, Berlin, 2005)
11. J. Bourjaily, H. Thomas, What is the Amplituhedron? Not. AMS **65**, 167–169 (2018)
12. A. Buch, A Littlewood–Richardson rule for the K-theory of Grassmannians. Acta Math. **189**(1), 37–78 (2002)
13. A. Buch, A. Kresch, K. Purbhoo, H. Tamvakis, The puzzle conjecture for the cohomology of two-step flag manifolds. J. Alg. Comb. **44**(4), 973–1007 (2016)
14. L. Chen, Quantum cohomology of flag manifolds. Adv. Math. **174**(1), 1–34 (2003)
15. I. Coskun, A Littlewood–Richardson rule for partial flag varieties, preprint (under revision), http://www.homepages.math.uic.edu/~coskun/newpartial.pdf
16. I. Coskun, A Littlewood–Richardson rule for two-step flag varieties. Invent. Math. **176**, 325 (2009)
17. D. Cox, J. Little, D. O'Shea, *Ideals, Varieties, and Algorithms*, vol. 2 (Springer, Berlin, 1991)
18. J.S. Frame, G.de B. Robinson, R.M. Thrall, The hook graphs of the symmetric group. Can. J. Math. **6**, 316–325 (1954)
19. S. Fomin, S. Gelfand, Quantum Schubert polynomials. J. Amer. Math. Soc. **10**, 565–596 (1997)
20. S. Fomin, A. Kirillov, Combinatorial B_n-analogues of Schubert polynomials. Trans. Am. Math. Soc. **348**(9) (1996)
21. W. Fulton, *Intersection Theory*, vol. 2 (Springer, Berlin, 1998)
22. W. Fulton, *Young Tableaux, with Applications to Representation Theory and Geometry* (Cambridge University Press, Cambridge, 1997)
23. A.M. Garsia, C. Procesi, On certain graded S_n-modules and the q-Kostka polynomials. Adv. Math. **94**(1), 82–138 (1992)
24. V. Gasharov, V. Reiner, Cohomology of smooth Schubert varieties in partial flag manifolds. J. Lond. Math. Soc. **66**(3), 550–562 (2002)
25. M. Gillespie, J. Levinson, K. Purbhoo, A crystal-like structure on shifted tableaux (2017), arxiv:1706.09969
26. R. Green, *Combinatorics of Minuscule Representations (Cambridge Tracts in Mathematics 199)* (Cambridge University Press, Cambridge, 2013)
27. A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, 2001)
28. S. Kleiman, D. Laskov, Schubert calculus. Am. Math. Mon. **79**(10), 1061–1082 (1972)
29. A. Knutson, T. Tao, C. Woodward, The honeycomb model of $GL_n(\mathbb{C})$ tensor products II: puzzles determine facets of the Littlewood–Richardson cone. J. Am. Math. Soc. **17**, 19–48 (2004)
30. A. Knutson, P. Zinn-Justin, Schubert, puzzles and integrability I: invariant trilinear forms (2017), arxiv:1706.10019v4

31. E.R. Kolchin, Algebraic matric groups and the Picard–Vessiot theory of homogeneous linear ordinary differential equations. Ann. Math. Second Ser. **49**, 1–42 (1948)
32. M. Kontsevich, Yu. Manin, Gromov–Witten classes, quantum cohomology, and enumerative geometry. Commun. Math. Phys. **164**, 525–562 (1994)
33. *Kac-Moody Groups, Their Flag Varieties and Representation Theory*, vol. 204, Progress in Mathematics (Birkhäuser, Basel, 2002)
34. T. Lam, L. Lapointe, J. Morse, A. Schilling, M. Shimozono, M. Zabrocki, *k-Schur Functions and Affine Schubert Calculus* (Springer, Berlin, 2014)
35. V. Lakshmibai, B. Sandhya, Criterion for smoothness of Schubert varieties in $SL(n)/B$. Proc. Indian Acad. Sci. Math. Sci. **100**, 45 (1990)
36. A. Lascoux, M.-P. Schützenberger, Polynômes de Schubert, C. R. Acad. Sci. Paris Sér. I Math. **294**(13), 447–450 (1982)
37. S.J. Lee, Combinatorial description of the cohomology of the affine flag variety. Trans. Am. Math. Soc. arXiv:1506.02390 (to appear)
38. I. Macdonald, *Symmetric Functions and Hall Polynomials* (Oxford University Press, Oxford, 1979)
39. L. Manivel, *Symmetric Functions, Schubert Polynomials, and Degeneracy Loci*, American Mathematical Society (2001)
40. E. Mukhin, V. Tarasov, A. Varchenko, Schubert calculus and representations of the general linear group. J. Am. Math. Soc. **22**(4), 909–940 (2009)
41. O. Pechenik, A. Yong, Equivariant K-theory of Grassmannians. Forum Math. Pi **5**, 1–128 (2017)
42. O. Pechenik, A. Yong, Equivariant K-theory of Grassmannians II: the Knutson–Vakil conjecture. Compos. Math. **153**, 667–677 (2017)
43. A. Postnikov, D. Speyer, L. Williams, Matching polytopes, toric geometry, and the totally non-negative Grassmannian. J. Alg. Comb. **30**, 173–191 (2009)
44. P. Pragacz, Algebro-geometric applications of Schur S- and Q-polynomials, in *Topics in Invariant Theory*, vol. 1478, Springer Lecture Notes in Mathematics, ed. by M.P. Malliavin (Springer, Berlin, 1991), pp. 130–191
45. F. Ronga, Schubert calculus according to Schubert (2006), arxiv:0608784
46. Y. Ruan, G. Tian, Mathematical theory of quantum cohomology. J. Differ. Geom. **42**(2), 259–367 (1995)
47. K. Ryan, On Schubert varieties in the flag manifold of $SL(n, \mathbb{C})$. Math. Ann. **276**, 205–224 (1987)
48. *The Symmetric Group*, vol. 2 (Springer, New York, 2001)
49. H. Schubert, *Kalkül der abzählende Geometrie* (Teubner Verlag, Leipzig, 1789.)
50. F. Sottile, Frontiers of reality in Schubert calculus. Bull. AMS **47**(1), 31–71 (2010)
51. R. Stanley, *Enumerative Combinatorics*, vol. 2 (Cambridge University Press, Cambridge, 1999)
52. Adv. Math. Shifted tableaux and the projective representations of the symmetric group. **74**(1), 87–134 (1989)
53. T. Tajakka, Cohomology of the Grassmannian. Master's thesis, Aalto University, 2015
54. H. Thomas, A. Yong, A combinatorial rule for (co)minuscule Schubert calculus. Adv. Math. **222**(2), 596–620 (2009)
55. J. Tymoczko, Decomposing Hessenberg varieties over classical groups. Ph.D. thesis, 2010, arxiv:0211226
56. R. Vakil, A geometric Littlewood–Richardson rule. Ann. Math. **164**(2), 371–422 (2006)
57. J.S. Wolper, A combinatorial approach to the singularities of Schubert varieties. Adv. Math. **76**, 184–193 (1989)

Combinatorics of the Diagonal Harmonics

Angela Hicks

Abstract The Shuffle Theorem, recently proven by Carlsson and Mellit, states that the bigraded Frobenius characteristic of the diagonal harmonics is equal to a weighted sum of parking functions. In this introduction to the topic, we discuss the theorem and connections between it and the well-known Macdonald polynomials. Furthermore, we describe important combinatorial bijections which imply various restatements of the theorem and play an important role in its proof. Finally, we briefly discuss the proof and describe various generalizations of the theorem.

The diagonal harmonics are a simply defined vector space of polynomials in the variables $X_n = \{x_1, \cdots, x_n\}$ and $Y_n = \{y_1, \cdots, y_n\}$:

Definition (Diagonal harmonics)**.**

$$DH_n = \{f(X_n, Y_n) \in \mathbb{C}[X_n, Y_n] : \sum_{i=1^n} \partial_{x_i}^r \partial_{y_i}^s f(X_n, Y_n) = 0 \text{ for all } r, s \geq 0,\ r + s > 0\},$$

with a natural $\mathfrak{S}_n$ action which permutes the X_n and Y_n variables simultaneously by acting on their indices. As a vector space, DH_n has proven to be a nontrivial puzzle to mathematicians, with questions like "What is its dimension?" inspiring new tools in algebraic combinatorics. Haiman [33] proved, using tools from algebraic geometry, that

$$\dim(DH_n) = (n+1)^{(n-1)}.$$

This suggested that perhaps one could supplement complex calculations about DH_n by finding a concrete family of combinatorial objects of size $(n+1)^{(n-1)}$ which can be used as a combinatorial model for algebraic properties of DH_n. The Shuffle Theorem, recently proven by Carlsson and Mellit [12], does exactly this, tying the combinatorics of parking functions to DH_n in a precise way.

A. Hicks (✉)
Lehigh University, Bethlehem, PA, USA
e-mail: anh316@lehigh.edu

H. Barcelo et al. (eds.), *Recent Trends in Algebraic Combinatorics*,
Association for Women in Mathematics Series 16,
https://doi.org/10.1007/978-3-030-05141-9_5

In what follows, we begin with a brief discussion of parking functions and their early appearance in the literature, followed by additional information on Macdonald polynomials meant to give context for the Shuffle Theorem and why it has been of historic interest. In Sect. 2, we carefully state the Shuffle Theorem and review the relevant definitions required to understand it. In Sect. 3, we describe from several perspectives a couple of important maps on the parking functions which lead to combinatorial restatements of the Shuffle Theorem. These maps, which occur at key points in the proof of the Shuffle Theorem and its generalization, have several dramatically different characterizations within the literature, and its our hope that by giving them and their various descriptions in one place, we'll make the connection between the older and newer literature more accessible. In Sect. 4, we give some necessarily vague (limited by our abbreviated text) indication of the final proof of the Shuffle Theorem. Finally, in Sect. 5 we give information about how the Shuffle Theorem has been generalized. We would be remiss if we did not immediately mention two well-written texts in this area which cover most of these topics and many more in much greater detail; both Haglund [28] and Bergeron [7] are excellent starting places for additional details, although they were written before the proof of the Shuffle Theorem was completed.

1 Background

Parking functions show up in a number of guises, from their initial inception as an idealized data storage method popular in theoretical computer science, to probability where their various statistics have ties to surprising distributions (see Yan [59] for a summary), and to algebraic combinatorics (our primary interest), where they are used to explore the representation theory of the diagonal harmonics via the Shuffle Theorem.

In their simplest form, parking functions are a subset of $(\mathbb{Z}/n\mathbb{Z})^n$; here we follow the convention that we use $\{1, 2, \cdots, n\}$ (but not 0) for the names of the equivalence classes. Think of

$$\pi = (\pi_1, \ldots, \pi_n)$$

as giving the preferences of a list of drivers, so that the ith driver wants to park in space π_i. The drivers proceed (one by one) to park in n spaces, with the first driver parking in his preferred space π_1. The ith driver does the same, unless his preferred space π_i is already occupied by a previous driver, in which case he proceeds to the next open space. If we imagine that there are exactly n spaces, π is a parking function if and only if every driver is able to park. We call the set of parking functions on n spaces PF_n.

It's easy to see that if any two drivers want to park in space n, π will fail to be a parking function since only $n - 2$ drivers will consider parking in the first $n - 1$ spaces, and thus one such space will be empty. Expanding the idea slightly, it must be that if π is in PF_n, for all $1 \leq j \leq n$,

$$\#\{i : \pi_i \leq j\} \geq j.$$

Remarkably this is a necessary and sufficient condition for $\pi \in (\mathbb{Z}/n\mathbb{Z})^n$ to be a parking function, and one often sees parking functions defined this way directly:

Definition (Parking Function)**.** $\pi = (\pi_1, \pi_2, \cdots, \pi_n)$ in $(\mathbb{Z}/n\mathbb{Z})^n$ is a parking function (i.e., is in PF_n) if for $1 \leq j \leq n$, $\#\{i : \pi_i \leq j\} \geq j$.

Thus, for example, (5, 2, 1, 4, 2) is a parking function, but (5, 2, 1, 5, 2) is not.

1.1 Parking Functions in Enumerative Combinatorics

Pyke [47] was the first to define parking functions, using somewhat different motivation and language than our definition above. The language of cars on a one way street is due to Konheim and Weiss [38] who described them while working for IBM Research. Their colorful (if not politically correct) story of "dutiful" husbands driving down a one way street until their "cantankerous" wives order them to park was the beginning of an analysis of a basic hashing technique, where information is stored on a disk at a specified location, unless that location is already in use. Computer science literature refers to this as "linear probing in a successful search" (as, for example, in Knuth [37]), and in this language, there is a lot known about the enumeration of parking functions according to various statistics. Pollak (as relayed by Foata and Riordan [13]) gives the best-known proof of the size of PF_n:

Theorem 1. $|PF_n| = (n+1)^{n-1}$.

Proof. Consider a street with $n+1$ spaces (so π is in $(\mathbb{Z}/(n+1)\mathbb{Z})^n$), where those spaces are set out along a circular street, so that if the ith driver finds parking spots π_i through spots $n+1$ occupied, he circles back to the beginning of the street to continue his hunt at the first spot. In this scenario, every π "parks" and since there is one more space than the number of cars, we have exactly one space that remains empty. Moreover, if we consider the resulting sequence of parked cars when we start with the preference π and when we start with the preference $\pi + (1, \ldots, 1) \in (\mathbb{Z}/(n+1)\mathbb{Z})^n$, the latter is a rotation of the former by exactly one space (with the empty space rotating similarly). Thus for any π,

$$\{\pi + k(1, \ldots, 1) : k \in \mathbb{Z}/(n+1)\mathbb{Z}\}$$

can be thought of as an equivalence class of functions whose resulting parked cars are the same up to rotation; it is clearly size $(n+1)$. Moreover, π in this setup is a parking function if and only if when drivers park according to π, the empty parking space is at $n+1$. Thus the parking functions make particularly nice representatives of each class and there must be $\frac{(n+1)^n}{(n+1)} = (n+1)^{n-1}$ parking functions as desired. $\square$

A number of other families of objects are in bijection with parking functions: the OEIS entry for $a_n = (n+1)^{n-1}$ includes, in addition to parking functions, spanning trees on the labeled complete graph on n vertices, the number of edge-labeled rooted trees on n nodes, the number of ways of expressing an n cycle as an a product of $n-1$ transpositions, as well as a number of others [54].

1.2 Symmetric Function Background

Parking functions play an interesting role in symmetric function theory and $\mathfrak{S}_n$ representation theory. Statistics on parking functions historically led to the development of a combinatorial formula for one of the important families of symmetric function bases: the Macdonald polynomials.

In the following, we use Λ^n for the vector space generated by homogeneous symmetric functions of degree n over a field $\mathbb{K}$, $\{m_\lambda\}_{\lambda\vdash n}$ for the monomial basis, $\{e_\lambda\}_{\lambda\vdash n}$ for the elementary basis, $\{h_\lambda\}_{\lambda\vdash n}$ for the homogeneous basis, $\{p_\lambda\}_{\lambda\vdash n}$ for the power sum basis, and $\{s_\lambda\}_{\lambda\vdash n}$ for the Schur basis. Stanley [55] and Sagan [51] contain definitions of all these bases for the unfamiliar reader.

We use $\trianglelefteq$ for dominance order on partitions, so that $\lambda \trianglelefteq \mu$ if for all i, $\lambda_1 + \cdots + \lambda_i \leq \mu_1 + \cdots + \mu_i$. If we use $m_i(\lambda)$ for the multiplicity of i in λ (i.e., $m_i(\lambda) = \#\{j : \lambda_j = i\}$), then recall the Hall inner product can be defined by:

$$\langle p_\lambda, p_\mu \rangle = z_\lambda \mathbb{1}_{\lambda=\mu},$$

where $z_\lambda = \prod_i i^{m_i(\lambda)} m_i(\lambda)!$ and $\mathbb{1}_X$ is the standard indicator function that is 1 if X is true and 0 if X is false. With these definitions in hand, then recall that one can uniquely define the Schur functions $\{s_\lambda\}_{\lambda\vdash n}$ by their following properties:

(1) (Orthonormality.) $\langle s_\lambda, s_\mu \rangle = \mathbb{1}_{\lambda=\mu}$.
(2) (Triangularity.) $s_\lambda = \sum_{\mu \trianglelefteq \lambda} K_{\lambda,\mu} m_\lambda$ where $K_{\lambda,\lambda} = 1$.

Recall that Schur function expansions are important in our understanding of the image of the Frobenius characteristic map. For a given class function ϕ of $\mathfrak{S}_n$, recall that

$$\mathrm{Fchar}(\phi) := \frac{1}{n!} \sum_\mu z_\mu^{-1} \phi(\mu) p_\mu.$$

The irreducible representations of $\mathfrak{S}_n$ are indexed by partitions, and the Frobenius map takes such an irreducible representation to a Schur function indexed by the same partition. Thus every $\mathfrak{S}_n$ module has an associated Schur positive polynomial, and in reverse, when a polynomial is Schur positive, it is reasonable to search for a natural associated $\mathfrak{S}_n$ module. Sagan [51] gives a nice introduction to these facts for the unfamiliar reader.

1.3 Macdonald Polynomials

Beyond the above five bases for Λ^n, a basis of particular importance is $\{P_\lambda[X; q, t]\}_{\lambda \vdash n}$, which we now call the Macdonald polynomials. (We now work in $\mathbb{K}[q, t]$.) Macdonald [44] originally defined them as a q, t analogue of the Schur basis. In particular, he began with a q, t analogue of the Hall inner product:

$$\langle p_\lambda, p_\mu \rangle_{q,t} = z_\lambda \prod_i \frac{(1 - q^{\lambda_i})}{(1 - t^{\lambda_i})} \mathbb{1}_{\lambda=\mu}.$$

Then $P_\lambda[X; q, t]$ is uniquely defined by:

(1) (Orthogonality.) $\langle P_\lambda[X; q, t], P_\mu[X; q, t] \rangle_{q,t} = c_\lambda \mathbb{1}_{\lambda=\mu}$.
(2) (Triangularity.)

$$P_\lambda[X; q, t] = \sum_{\mu \trianglelefteq \lambda} d_{\lambda,\mu}(q, t) m_\mu,$$

where $d_{\lambda,\lambda}(q, t) = 1$.

Replacing t or q in $P_\lambda[X; q, t]$ with various integer values (usually 0, 1, or a limit) gives a number of famous symmetric function bases, from the Schur functions themselves (at $t = q = 0$) and monomials ($t = 1$) to the Hall–Littlewood, the Jack, the Askey–Wilson, and the Koornwinder polynomials. In short, Macdonald polynomials encode information about most of the important bases of the symmetric functions.

Although $P_\lambda[X; q, t]$ is referred to as a Macdonald *polynomial*, strictly speaking it is a symmetric function, with coefficients which are rational functions in q and t; Macdonald also defined $J_\lambda[X; q, t] = c_\lambda(q, t) P_\lambda[X; q, t]$ which is a different rescaled version of $P_\lambda[X; q, t]$ using a combinatorially defined constant $c_\lambda(q, t)$, chosen such that $J_\lambda[X; q, t]$ has coefficients which are polynomial in q and t. (See Macdonald [44, 8.1].) Macdonald's Schur Positivity Conjecture (see Macdonald [44, 8.18]) stated that $J_\lambda[X; q, t]$ was "almost" Schur positive: expressing $J_\lambda[X; q, t]$ in terms of a closely related t-analogue of the Schurs $\{s_\mu(x; t)\}_{\mu \vdash n}$ gave coefficients that are polynomials in q and t with positive integral coefficients. More precisely, $\{s_\mu[X; t]\}_{\mu \vdash n}$ differs from $\{s_\mu\}_{\mu \vdash n}$ by a simple homomorphism: let ϕ be the multiplicative homomorphism that for all nonnegative integers, k sends the power sum p_k to $p_k(1 - t^k)$. Then

$$s_\lambda[X; t] := \phi(s_\lambda) = s_\lambda[X(1 - t)],$$

where the last term gives the same expression using plethystic notation. See Loehr and Remmel [43] for complete details on the notation.

Garsia suggested several different manipulations of the Macdonald polynomials, applying the inverse homomorphism (and flipping the powers of t– sending $t \to \frac{1}{t}$ and multiplying by a power of t) defining two new families of polynomials.

$$H_\lambda[X;q,t] := J_\lambda\left[\frac{X}{1-t};q,t\right]$$

and

$$\tilde{H}_\lambda[X;q,t] := t^{n(\mu)}H_\lambda[X;q,1/t],$$

where $n(\mu)$ is just the highest power of t in $H_\lambda[X;q,t]$. All the above-mentioned families of polynomials from the $P_\lambda[X;q,t]$ to the $\tilde{H}_\lambda[X;q,t]$ are referred to by various authors as "the Macdonald polynomials," although their chosen notation—using P, J or $\tilde{H}$—generally differentiates them.

The modified Macdonald polynomials are in fact Schur positive, as Macdonald's Schur Positivity Conjecture. Garsia and Haiman [17] gave a combinatorially defined polynomial Δ_λ for each partition of λ of n. In particular, for a Ferrer's diagram of the partition λ drawn in French notation, with its southwest corner at the origin, Garsia and Haiman take the set $\{(p_1,q_1),\cdots,(p_n,q_n)\}$ of southwest corners of each cell in the diagram. (Thus in particular, (0, 0) will correspond to the bottom left cell.) Then, they define

$$\Delta_\lambda = \det \|x_i^{p_j}y_i^{q_j}\|_{i,j=1,\cdots,n}.$$

See Fig. 1 for an example when $\lambda = (3,2)$. Garsia and Haiman then considered the linear span of all the partial derivatives of the Δ_λ, which we will denote $\mathcal{L}[\partial_x\partial_y\Delta_\lambda]$. Moreover, they showed that if for all λ, the dimension of $\mathcal{L}[\partial_x\partial_y\Delta_\lambda]$ is $n!$, then the Frobenius image of $\mathcal{L}[\partial_x\partial_y\Delta_\lambda]$ is $\tilde{H}_\lambda[X;q,t]$, and thus, the modified Macdonald polynomials are Schur positive. The statement about the dimension is the well-known $n!$ Conjecture, which was proved almost a decade later in Haiman [32]. His proof thus implied that the coefficient of q^it^j in $\tilde{H}_\lambda[X;q,t]$ is the Frobenius character of the subspace of $\mathcal{L}[\partial_x\partial_y\Delta_\lambda]$ of total degree i in the x variables and j in the y variables, and moreover, that Macdonald's Schur Positivity Conjecture is true. While Haglund et al. [30] gives a combinatorial formula (as opposed to an indirect definition via triangularities) for the modified Macdonald polynomials, there is great interest in a general formula the modified Macdonald–Kostka coefficients

$$\tilde{K}_{\mu,\lambda}(q,t) := \left\langle \tilde{H}_\lambda[X;q,t], s_\mu \right\rangle.$$

At the time of publication, the most complete results in this direction are presented in Assaf [4].

2 The Diagonal Harmonics

A closely related space to $\mathcal{L}[\partial_x\partial_y\Delta_\lambda]$ is the diagonal harmonics; recall that as mentioned above:

Definition (Diagonal harmonics)**.**

(0,1)	(1,1)	
(0,0)	(1,0)	(2,0)

$$\Delta_{3,2} = \det \left(\begin{bmatrix} x_1^0y_1^0 & x_1^1y_1^0 & x_1^2y_1^0 & x_1^0y_1^1 & x_1^1y_1^1 \\ x_2^0y_2^0 & x_2^1y_2^0 & x_2^2y_2^0 & x_2^0y_2^1 & x_2^1y_2^1 \\ x_3^0y_3^0 & x_3^1y_3^0 & x_3^2y_3^0 & x_3^0y_3^1 & x_3^1y_3^1 \\ x_4^0y_4^0 & x_4^1y_4^0 & x_4^2y_4^0 & x_4^0y_4^1 & x_4^1y_4^1 \\ x_5^0y_5^0 & x_5^1y_5^0 & x_5^2y_5^0 & x_5^0y_5^1 & x_5^1y_5^1 \end{bmatrix} \right)$$

Fig. 1 Computing $\Delta_{3,2}$

$$DH_n = \{f(X_n, Y_n) \in \mathbb{C}[X_n, Y_n] : \sum_{i=1^n} \partial_{x_i}^r \partial_{y_i}^s f(X_n, Y_n) = 0 \text{ for all } r, s \geq 0,\ r+s > 0\}.$$

Theorem 2. *If $\mu \vdash n$, then $L[\partial x \partial y \Delta_\mu]$ is a subspace of DH_n.*

Proof. Let M_μ be the matrix whose determinant is Δ_μ, and let $(M)^{i,j}$ denote the (i,j)th minor of M_μ. Let $\frac{\partial}{\partial_k(a,b)} = \frac{\partial}{\partial x_k^a}\frac{\partial}{\partial y_k^b}$. $\frac{\partial}{\partial_k(a,b)}\Delta_\mu$ is clearly 0 if (a,b) is not a cell which occurs in μ, so we assume $(a,b) = (p_l, q_l)$ for some l in $\{1, \cdots, n\}$.

$$\begin{aligned} \sum_{k=1}^n \frac{\partial}{\partial_k(p_l,q_l)} \Delta_\mu &= \sum_{k=1}^n \frac{\partial}{\partial_k(p_l,q_l)} \det \|x_i^{p_j} y_i^{q_j}\|_{i,j=1,\cdots,n} \\ &= \sum_{k=1}^n \frac{\partial}{\partial_k(p_l,q_l)} \sum_{j=1}^n (-1)^{j+k} x_k^{p_j} y_k^{q_j} (M_\mu)^{j,k} \\ &= \sum_{j=1}^n \sum_{k=1}^n (-1)^{j+k} \left(\frac{\partial}{\partial_k(p_l,q_l)} x_k^{p_j} y_k^{q_j} \right) (M_\mu)^{j,k} \end{aligned}$$

For a fixed j, the final summation is also the determinant of a matrix calculated using expansion by minors, in particular M_μ with the jth column replaced by the given partial derivatives. Moreover, the jth column is either identically 0 or a scalar multiple of another column in M_μ corresponding to a cell in the Ferrer's diagram of μ which is south and west. Thus for every j, the determinant is 0.

Haiman [33] computed the dimension of both DH_n and $\mathcal{L}[\partial_x \partial_y \Delta_\lambda]$ using similar techniques from algebraic geometry, in particular employing the Hilbert scheme of points in the plane, using a method of approach first sketched by Procesi years before. As will be discussed in more detail below, eigenoperators of the modified Macdonald polynomials play a key role in the symmetric function theory used to study the larger space of diagonal harmonics.

The bigraded Frobenius characteristic of the diagonal harmonics has had a conjectured combinatorial formula since the 1990s (before any combinatorial formula was suggested for the Macdonald polynomials), although the proof wasn't completed until two decades later. Let $DH_n^{a,b}$ give the submodule of DH_n that is total degree a in the 'x' variables and b in the 'y' variables. Then we can define the bigraded

Frobenius characteristic as

$$DH_n[X;q,t] := \sum_{a,b} t^a q^b \,\mathrm{Fchar}\, DH_n^{a,b}.$$

The Shuffle Theorem, first conjectured by Haglund et al. [30], was proved recently by Carlsson and Mellit [12], building on the previous work of Haiman [33]:

Theorem 3 (The Shuffle Theorem)**.**

$$DH_n[X;q,t] = \sum_{\pi \in PF_n} t^{\mathrm{area}(\pi)} q^{\mathrm{dinv}(\pi)} F_{\mathrm{ides}(\pi)}. \tag{1}$$

In the following, we'll discuss the definitions of area, dinv, and ides. F_S is Gessel's fundamental quasisymmetric function. Fundamental quasisymmetric functions are often represented with the subscript S a composition of n, but as elsewhere in the literature on parking functions, it will be simpler to assume the subscript S is a subset of $\{1, 2, \cdots, n-1\}$ and not a composition of n. There is, of course, a standard bijection between the two, which can be employed to switch from one convention to the other.

It's worth noting that that even in the literature written after its proof, Theorem 3 is still often referred to as "The Shuffle Conjecture." More details are contained in Sect. 4 about its proof.

Researchers routinely picture parking functions more diagrammatically to compute the statistics appearing in the right side of the theorem. As first suggested by Adriano Garsia, parking functions are viewed in an $n \times n$ lattice with:

(1) A Dyck path—north and east steps from the bottom left to top right which stay (weakly) above the line $x = y$.
(2) Labels (or "cars")—the numbers 1 to n, each occurring exactly once to the right of a north step, so that they are increasing from bottom to top within a column.

To see that this definition matches our first definition, starting from such a diagram, define $\pi = (\pi_1, \cdots, \pi_n)$ such that $\pi_i = j$ if a j is in the ith column of the lattice diagram. The Dyck path condition is equivalent to the rule that we must have at least k cars less than or equal to k for all $1 < k < n$. The increasing column condition defines a fixed ordering of the cars within a given column, which all had the same preferred spot. (Without this or a similar an ordering, the map would not be bijective.) See Fig. 2 for an example.

Starting with a parking function represented in a lattice, we then have the following definition.

Definition (area)**.** The area of a parking function π is the number of complete squares between the Dyck path of π and the main diagonal and will be denoted area(π).

Example 1. The area of the parking function in Fig. 2 is (adding it by row) $0 + 0 + 1 + 2 + 1 + 0 + 1 + 0 + 0 = 5$.

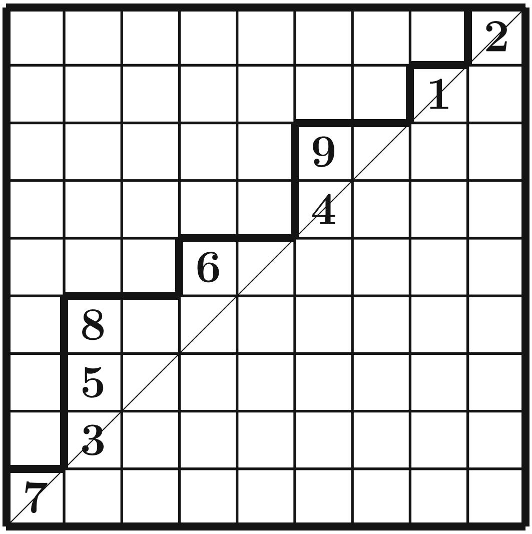


Fig. 2 A lattice diagram of the parking function (8, 9, 2, 6, 2, 4, 1, 2, 6) in PF_9 as drawn in a lattice diagram

We refer to a diagonal of the parking function as the cells along a line $y = x + b$ for some $0 \leq b < n$.

Definition (dinv)**.** A primary diagonal inversion is a pair of cars in the same diagonal, with the smaller further southwest. A secondary diagonal inversion is a pair of cars in two adjacent diagonals, with the smaller car in the lower diagonal (with lower y-intercept), but northeast of the larger car in the higher diagonal. The dinv of a parking function π is the number of primary and secondary diagonal inversions and will be denoted $\text{dinv}(\pi)$.

Example 2. In Fig. 2, the pair $\{5, 9\}$ is a primary diagonal inversion, but $\{1, 4\}$ is *not* a primary inversion since the larger car (4) is further southwest. Similarly, $\{4, 5\}$ makes a secondary inversion, but $\{8, 9\}$ does *not* (because of the relative positions of the smaller and larger car) nor would any choice of car number that could be placed in the cells currently containing $\{4, 8\}$ (since the cells aren't on adjacent diagonals) nor do $\{3, 9\}$ (where the smaller car on the lower diagonal is not northeast of the larger car.) The dinv of the parking function in Fig. 2 is 14, including the pairs: $\{\mathbf{3}, 4\}$, $\{\mathbf{1}, 2\}$, $\{1, \mathbf{9}\}$, $\{2, \mathbf{9}\}$, $\{1, \mathbf{6}\}$, $\{2, \mathbf{6}\}$,$\{\mathbf{6}, 9\}$,$\{4, \mathbf{6}\}$,$\{1, \mathbf{5}\}$,$\{2, \mathbf{5}\}$, $\{\mathbf{5}, 9\}$, $\{4, \mathbf{5}\}$, $\{\mathbf{5}, 6\}$, and $\{6, \mathbf{8}\}$. We've bolded the southwest car in each pair, for reasons which will be more apparent later.

If two cells are positioned so that they could create a diagonal inversion if the two numbers contained within them are in the "correct" order (as, for example, in Fig. 2 both the pair $\{3, 4\}$, which makes an diagonal inversion and $\{1, 4\}$, which does not), then the cells are called attacking.

Definition (word)**.** We read the word of a parking function by reading along diagonals from highest to lowest, within a diagonal reading from right to left.

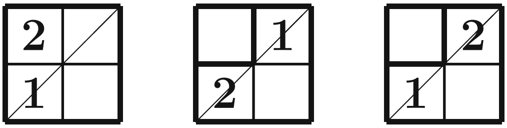


Fig. 3 These three parking functions have weights $tF_{\{1\}}$,$F_{\{\emptyset\}}$, and $qF_{\{1\}}$, respectively

In particular, note that the word of a parking function is a permutation in $\mathfrak{S}_n$ and thus has a well-defined inverse.

Definition (ides)**.** The i-descent set of a permutation is defined to be the descent set of the inverse of a permutation, that is, the set of i in a permutation such that $i+1$ occurs before i. The i-descent set of a parking function is the descent set of the inverse of the word of a parking function.

Example 3. The word of the parking function in Fig. 1 is $[1, 2, 3, 4, 5, 6, 7, 8, 9]$. It's i-descent set is $\{1, 3, 4, 5, 7\}$.

Example 4. There are three parking functions of size 2, pictured in Fig. 3. Computing their weights, we get that

$$DH_2[X; q, t] = tF_{\{1\}} + F_{\emptyset} + qF_{\{1\}} = (t + q)s_{(1,1)} + s_{(2)}.$$

Thus by the Shuffle Theorem, DH_2 contains a submodule of degree 0 that is isomorphic to the trivial representation of $\mathfrak{S}_2$ and two submodules, one of degree 1 in the x variables and one of degree 1 in the y variables, that are isomorphic to the signed representation.

While these statistics (particularly dinv) may not seem particularly natural, their discovery was important not just to the Shuffle Theorem, but also to the combinatorics of the Macdonald polynomials. As in the Shuffle Theorem, combinatorial interpretations of Macdonald Polynomials require two statistics, one of which (maj) is relatively natural, and another (inv) whose discovery was motivated by dinv in parking functions. Haglund [27] describes the process by which he first defined inv, which is a more complicated statistic than dinv, from dinv on parking functions.

2.1 Shuffles and the q, t Catalan

For convenience, we name the sum in the right-hand side of (1). Let

$$\Pi_n(X; q, t) := \sum_{\pi \in PF_n} t^{\mathrm{area}(\pi)} q^{\mathrm{dinv}(\pi)} F_{\mathrm{ides}(PF)}.$$

$\Pi_n(X; q, t)$ is a positive sum of LLT polynomials, which are Schur positive polynomials developed in Lascoux et al. [39] and Leclerc and Thibon [40]. While the

original definition of LLT polynomials (in terms of a statistic on ribbon tableaux) does not generally "look" very similar to parking functions, work in Schilling et al. [52], simplified independently by Haiman and Michelle Bylund, gives an equivalent definition which is used in Haglund et al. [30] to show that $\Pi_n(X; q, t)$ is Schur positive. Since in particular $\Pi_n(X; q, t)$ (as a sum of LLT polynomials) is symmetric, it makes sense to consider the Hall inner product of $\Pi_n(X; q, t)$ with $h_{k_1} \cdots h_{k_s}$. Let

$$Sh(k_1, \cdots, k_s) = \{1, \cdots, k_1\} \sqcup\!\sqcup \{k_1 + 1, \cdots, k_1 + k_2\} \sqcup\!\sqcup \cdots \sqcup\!\sqcup \{k_1 + \cdots k_{s-1} + 1, \cdots, k_1 + k_2 + \cdots + k_s\}$$

give the set of shuffles (words with $1, \cdots, k_1$ occurring relatively in increasing order, as are $\{k_1 + 1, \cdots, k_1 + k_2\}$ and so on). Then basic symmetric function theory gives:

$$\langle \Pi_n(X; q, t), h_{k_1} \cdots h_{k_s} \rangle = \sum_{\substack{\pi \in PF_n \\ \text{word}(\pi) \in Sh(k_1, \cdots, k_s)}} t^{\text{area}(\pi)} q^{\text{dinv}(\pi)}$$

and motivates the name of the *Shuffle* Theorem. Similarly,

$$\langle \Pi_n(X; q, t), e_n \rangle = \sum_{\substack{\pi \in PF_n \\ \text{word}(\pi) = (n, n-1, \cdots, 1)}} t^{\text{area}(\pi)} q^{\text{dinv}(\pi)}.$$

If we restrict ourselves to this case, since the word is completely determined, we can just leave the car numbers off entirely from the diagram; the result for each such parking function is a unique Dyck path. Note that this implicitly defines the dinv and area of a Dyck path, with the area being the sum of the cells below the path as before and the dinv simplifying, since the conditions on the relative size of the cars are always fulfilled, to the number of attacking cells. Since Dyck paths are counted by Catalan numbers, this motivates the name used to describe this case: the q, t-Catalan Theorem, which was first proved by Garsia and Haglund [16]. Historically, dinv (and before that bounce, as described below) was developed first on Dyck paths, not on parking functions. Since we describe the Shuffle Theorem here first, and then give the q, t- Catalan case as a specialization, our presentation here is in the reverse order of the historical development, where the Catalan case was generalized to the Shuffle Theorem.

3 Alternate Formulations of the Shuffle Conjecture: The ζ Map and the Γ Map

One map that is particularly important to the proof of the Shuffle Theorem is the ζ map, which shows $\Pi_n(X; q, t)$ can equivalently be computed as a sum over "diagonally labeled Dyck paths"—a variant on parking functions. A second map, the Γ map, which when restricted to Dyck paths agrees with the ζ map, proves a particular

symmetry of the parking functions and leads to a natural generalization of the shuffle conjecture.

3.1 The ζ Map on Dyck Paths

There are several equivalent formulations of the ζ map on Dyck paths. Perhaps the simplest (although not the original) definition comes from the language of sweep maps in Armstrong et al. [3]. Here, as elsewhere, we find it convenient to picture a Dyck path as not in an $n \times n$ grid, but an $n \times (n + 1)$ grid, with the Dyck path remaining above the diagonal from the bottom left corner to the top right (so that the diagonal is tilted just slightly clockwise to "break ties"). There are the same Catalan number of such paths in either grid, with "Dyck paths" in the $n \times (n + 1)$ grid being created from ones in the standard square grid by adding a final east step.

We will need an ordering on each of the individual north and east steps in a Dyck path. Such an ordering will be defined by studying the south end of a north step and the west end of an east step; that is, our ordering on the steps is based on the position of the first point we would touch each step if we were to trace along the path from bottom left to top right.

Draw π as a Dyck path in $n \times (n + 1)$ grid. Then create a new path $\zeta(\pi)$ by sweeping the main diagonal upward, and recording the new north and east steps as a new path as they (or rather their south and west endpoints) are encountered.

While this description does not make it "obvious," ζ is a bijection from the Dyck paths of size n to itself. Moreover, it has a predictable effect on the statistics of the Dyck paths; in particular if π has area a and dinv b, $\zeta(\pi)$ has area b and bounce a, where bounce is a third statistic on Dyck paths (viewed on the $n \times n$ grid.)

Definition. To compute the bounce of a Dyck path π, trace a new Dyck path, called the "bounce path," starting at the point $(0, 0)$ and moving upward:

(1) Turn right when you hit the west endpoint of any east step of π.
(2) Resume an upward path after intersecting the line $x = y$.

Repeat these two steps until the path goes from $(0, 0)$ to (n, n).

Finally, if the bounce path intersects the line $x = y$ at the points

$$(0, 0), (n - i_1, n - i_1), \cdots , (n - i_k, n - i_k), (n, n),$$

then by definition

$$\text{bounce}(\pi) := i_1 + \cdots + i_k.$$

Example 5. In Fig. 4, the Dyck path on the right has bounce $1 + 6 + 8 = 15$

Using this definition and D_n for the set of Dyck paths of size n, as related in Haglund and Loehr [29, pg.9], we thus have

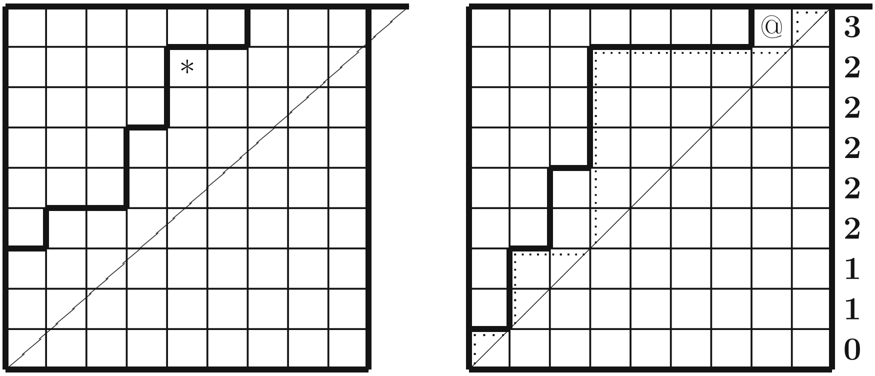


Fig. 4 π and $\zeta(\pi)$. The diagonal line is slightly tilted in the preimage as it occurs in the definition of the sweep map and emphasized the extra east step, but kept the diagonal line $x = y$ in the image, as it occurs in the included description of ζ^{-1}. The dashed line gives the bounce path. The numbers along the right correspond to after how many bounces the given row was crossed by the bounce path, but also indicate the diagonal containing the preimage of that north steps in the Dyck path on the left

Theorem 4 (Garsia, Haglund, Haiman)**.**

$$\sum_{\pi \in D_n} t^{\text{area}(\pi)} q^{\text{dinv}(\pi)} = \sum_{\pi \in D_n} t^{\text{bounce}(\pi)} q^{\text{area}(\pi)}. \tag{2}$$

A nice detail of the bijection is that area under the bounce path of $\zeta(\pi)$ corresponds to primary diagonal inversions in π and area under $\zeta(\pi)$ but above the bounce path corresponds to secondary inversions. Extending this observation slightly gives the inverse map, most easily stated as onto the set of parking functions with word $[n, n-1, \cdots, 1]$. (In the image, $\zeta(\pi)$, it's useful to consider adding the number from bottom to top to get a parking function. In $\pi = \zeta^{-1}(\zeta(\pi))$, we'll find the result has word, as read along the diagonals, $[n, n-1, \cdots, 1]$.) In particular, if the ith north step (number from the bottom) in $\zeta(\pi)$ was in a row crossed after the bounce path's jth return to the line $x = y$ (noting that the first step or more will thus be assigned the number 0, since the bounce path has not yet "returned"), then i was in the jth diagonal of π. (As a quick exercise to check for understanding, the interested reader may wish to check that the sum of such j must sum to the bounce of $\zeta(\pi)$, so that $\text{bounce}(\zeta(\pi)) = \text{area}(\pi)$, as the sum of the diagonals containing each car.) Since π originally corresponded to a Dyck path, the cars in any diagonal should be placed in increasing order from southwest to northeast. The resulting diagonal inversions in π (the primary diagonal inversions) account for area under the bounce path. We still need to account for the relative order of cars (and in particular which occurs in which row) in distinct diagonals. In particular, this is uniquely determined by the prescription that in π, car i is southwest of j smaller cars in the next lower (southeast) diagonal if in $\zeta(\pi)$, there are j cells in the ith row that were above the bounce path.

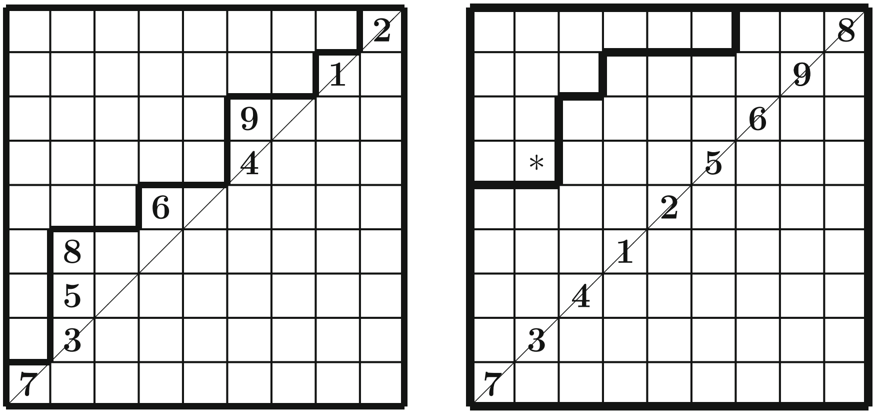


Fig. 5 π and $\zeta(\pi)$. Note that for any corner in $\zeta(\pi)$ like the one marked with a $*$, the car below the corner is smaller than the car to the right of the corner

Example 6. The '$*$' in π and the '@' in $\zeta(\pi)$ in Fig. 4 correspond to the same cell. (Note that the '$*$' is in the first cell that is read in the reading word of π and thus would be labeled $n = 9$, while the '@' occurs in the 9th row from the bottom.) The single north step in π northeast of '$*$' is exactly one diagonal lower, so it corresponds to a secondary inversion. It thus corresponds to a single cell above the bounce path in $\zeta(\pi)$ in the row containing '@', in particular the cell containing '@' itself.

3.2 ζ on the Parking Functions

There are two well-known alternate characterizations of $\Pi_n[X; q, t]$, both which extend ζ. We start with the summation tied to the bijection (again) called ζ in the literature. The image of ζ on parking functions is not a parking function. The image does not seem to have a standard name in the literature; we'll refer to it here as the diagonally labeled Dyck paths (DLD_n). The diagonally labeled Dyck paths are also represented in a lattice diagram, which again contains a Dyck path. Each Dyck path contains a permutation of n; this time the numbers (which we will here still call cars) are placed along the main diagonal. Rather than an increasing column condition, the only restriction on the values of the cars in DLD_n is based on certain corners in the Dyck path. In particular, if the underlying path contains an east step followed by a north step, the car below the east step must be smaller than the car to the right of the north step.

Example 7. See the right side of Fig. 5 for an example.

Definition. For $\pi \in PF_n$, we define $\zeta(\pi)$ as:

- First, if D is the Dyck path in π, create $\zeta(D)$.

- Add cars to the diagonal of $\zeta(D)$ from northeast to southwest, labeling the cells with the reading word of π.

The result is $\zeta(\pi)$.

Example 8. See Fig. 5 for an example.

The bijectivity of ζ on parking functions follows simply from the bijectivity of ζ on the Dyck paths. Moreover, we can again define statistics in the image which correspond to area and dinv in the preimage. Since area is constant on Dyck paths and we have defined ζ on Dyck paths to send area to bounce, it is clear that one such statistics should be the bounce of the path in $\zeta(\pi)$. The dinv of a parking function is mapped under ζ to the area of the result only when every attacking pair of cars in π creates a diagonal inversion, i.e., when the parking function has reading word $n, n-1, \cdots, 1$ and thus corresponds to a Dyck path. We need a new statistic, area' which is less than or equal to area.

Definition. The area' of $\pi \in DLD_n$ is the number of cells c between the Dyck path and the main diagonal such that the car below c is smaller than the car to the right of c.

Example 9. In the diagonally labeled path on the right of Fig. 5,

$$(\{3,4\},\{4,5\},\{4,6\},\{1,2\},\{1,5\},\{1,6\},\{1,9\},\{2,5\},\{2,6\},\{2,9\},\{5,6\},\{5,9\},\{6,9\},\{6,8\})$$

all contribute to the area' of 14. Note that these pairs are exactly the pairs that create dinv in the parking function on the left. The underlying Dyck path on the right has bounce 5, which corresponds to the area of 5 on the left.

Finally, we have the theorem:

Theorem 5 (Haglund and Loehr [29]).

$$\sum_{\pi \in PF_n} t^{\mathrm{area}(\pi)} q^{\mathrm{dinv}(\pi)} = \sum_{\pi \in DLD_n} t^{\mathrm{bounce}(\pi)} q^{\mathrm{area}'(\pi)}. \tag{3}$$

3.3 Γ on the Parking Functions

Unlike ζ, Γ is a map from PF_n to itself. This generalization leads naturally to rational parking functions and suggested a parallel story to the diagonal harmonics in other types. We'll describe it two ways, once as a combinatorial generalization of ζ on Dyck paths as described above and once as a composition of two bijections to a set of intermediate objects (a particular subset of affine permutations).

3.3.1 A Direct Combinatorial Description

Since all the attacking cells in a Dyck path create diagonal inversions but only some of them create diagonal inversions in a parking function, it's reasonable to suspect there might be a generalization of bounce that generally lowers the statistic depending on the location of cars and is equidistributed with area of the parking functions. In fact, such a statistic exists and is referred to in the literature as pmaj. While it can no longer be pictured using the geometry of a bouncing ball, when $\pi \in PF_n$ has car labels increasing from the bottom row to the top row, $\mathrm{pmaj}(\pi)$ is the same as the bounce of the underlying Dyck path. (As remarked previously, a close look at the inverse map described in Sect. 3.1 shows that we were implicitly injecting the Dyck paths into the parking functions in two ways: in the preimage of ζ, we considered them as parking functions with word $[n, n-1, \cdots, 1]$ but in the image, we considered parking functions with cars increasing from the bottom row to the top.) One way we might have equivalently described bounce is by a recursive procedure, starting by assigning any north step/car in the first column bounce 0. We then assigned bounce to cars in increasingly higher amounts. After assigning bounce i, by returning to the line $x = y$, we count the number of cars already assigned bounce i or less. If that number is j, we then assigned bounce $i+1$ to cars which are:

- in the $j+1$th column or less and
- have not yet been assigned bounce.

We then repeat the process, this time assigning bounce $i+2$. Pmaj can be assigned similarly, but we should expect that some cars will be assigned a pmaj earlier in the procedure (and thus a lower pmaj) on the basis of their value.

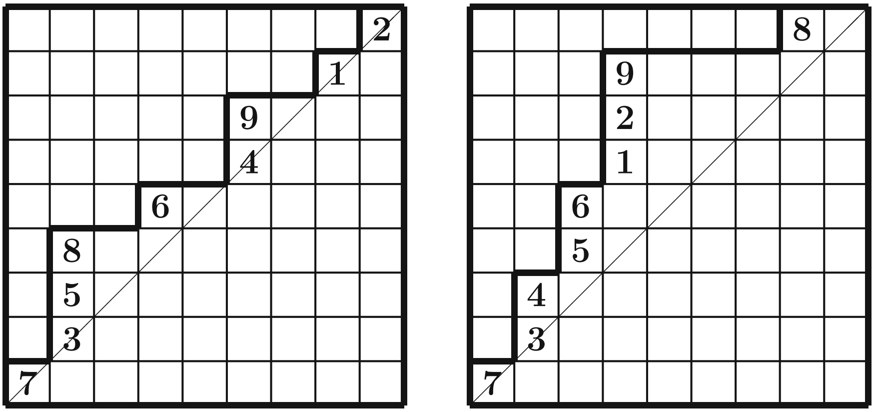
Formally, for a parking function π, the procedure for assigning pmaj is:

- Assign every car in the first column a pmaj of 0. Set p (which we'll use to keep track of the current pmaj we assign) and k (the number of cars assigned a pmaj of at least $p-1$) to 0.
- For every car c not yet assigned pmaj, count the number of larger cars already assigned pmaj exactly p, call it j. If c is in column $k+j+1$ or less, assign c the pmaj p. (It's best to work with the cars c in decreasing order of size, although one can equivalently repeat this step until no new cars are able to be added.)
- Increase p to $p+1$, and increase k to include the number of cars added in the previous step.
- Repeat (2) and (3) until all the cars are assigned a pmaj.

We will use $\mathrm{pmaj}(c, \pi)$ for the pmaj of the car c in the parking function π.

Definition (pmaj)**.** The pmaj of a parking function π ($\mathrm{pmaj}(\pi)$) is the sum of the pmaj assigned to each car.

Example 10. Consider the parking function on the **right** of Fig. 6. Car 7 is immediately assigned a pmaj of 0 since it's in the first column. Cars 9 and 8 are larger than 7 and so are not assigned pmaj (because they aren't in the $(0+0+1)$st or first column). Similarly, 6 and 5 aren't in the $(0+1+1)$th or second column. However,



car	1	2	3	4	5	6	7	8	9	car
area	0	0	0	0	1	1	0	2	1	pmaj
dinv	1	0	1	0	5	4	0	1	2	area

Fig. 6 π (repeated here from Fig. 2) and $\Gamma(\pi)$

4 and 3 are both in the 2nd column (and less than 7), so they are both assigned a pmaj of 0. Next, 3, 4, 7 (which have already been assigned pmaj this round) are all greater than 1 and 2, which are in the 4th column, so since $0 + 3 + 1 \geq 4$ and $0 + 4 + 1 \geq 4$, 2 and 1, respectively, are assigned pmaj 0. This ends the cars with pmaj 0.

Next, we've assigned five cars pmaj 0, so any unassigned car in the 6th column or less has pmaj 1, including cars 5, 6, and 9. Moreover, once car 9 is assigned a pmaj of 1, we compare the sum of the five previously assigned cars and the one larger car than 8 (the 9) which was recently assigned a pmaj of 1. Since $5 + 1 + 1 < 8$, we cannot assign 8 a pmaj of 1.

In a final assignment, we may use the eight cars assigned pmaj one or zero to justify a pmaj of 2 for car 8.

The observant reader may notice that the diagonal containing c in the parking function on the left of Fig. 6 is in every case the pmaj of c in $\Gamma(\pi)$. This is not a coincidence, but rather the starting place of the map we'll use to define Γ. We've implicitly assigned pmaj to each car; we'll do the same for area (in two ways, once for the preimages and once for the images) and dinv. The resulting assignments will be equidistributed and used to define the bijection.

For π in the preimage of Γ, first assign the area of car c ($\text{area}(c, \pi)$) as the number of complete squares weakly to the right of c and left of the line $x = y$ (i.e., the diagonal of the car c). Next, we must decide which of the two cars in a dinv pair should be assigned the diagonal inversion. We will find it useful to assign the dinv of a car c ($\text{dinv}(c, \pi)$) as the number of diagonal inversion pairs containing c and in which c is the further southwest of the pair. In particular, this means that a primary

diagonal inversion is assigned to the smaller of two cars, but a secondary inversion is assigned to the larger car.

Example 11. In Example 2, we've bolded the cars which are being assigned dinv in each pair of diagonal inversions in the left of Fig. 6.

For $\Gamma(\pi)$, we've already described the assignment of pmaj. Although the area of the parking function is defined the same way on each side, we need to assign area to individual cars differently in the image. In particular, the area of a particular car is closely tied to our assignment of pmaj and how many columns right c can move and still have the same pmaj. In particular, here we assign

$$\begin{aligned}&\mathrm{area_p}(c, \Gamma(\pi))\\&\quad = \#\{d \mid \mathrm{pmaj}(d, \Gamma(\pi)) < \mathrm{pmaj}(c, \Gamma(\pi)) + \#\{d > c \mid \mathrm{pmaj}(d, \Gamma(\pi)) = \mathrm{pmaj}(c, \Gamma(\pi))\}\\&\qquad\qquad - \mathrm{column}(c, \Gamma(\pi)) + 1.\end{aligned}$$

where $\mathrm{column}(c, \Gamma(\pi))$ gives the column containing c (as usual numbering from left to right). Then

$$\mathrm{area}(\Gamma(\pi)) = \sum_{c=1}^{n} \mathrm{area_p}(c, \Gamma(\pi)),$$

since if we sum over cars first sorting by their pmaj (small to large) then by car (large to small), we will find $\#\{d \mid \mathrm{pmaj}(d, \Gamma(\pi)) < \mathrm{pmaj}(c, \Gamma(\pi))\} + \#\{d > c \mid \mathrm{pmaj}(d, \Gamma(\pi)) = \mathrm{pmaj}(c, \Gamma(\pi))\}$ gives us the sequence $0, 1, 2, \cdots, n-1$ which sums to the number of cells strictly above the main diagonal in the lattice. Meanwhile $-\mathrm{column}(c, \Gamma(\pi)) + 1$, summed over all cars c, is the number of cells above the Dyck path (usually called the coarea), so the sum of all terms is the area between the path and the diagonal as claimed.

Recursive procedures detailed in Haglund and Loehr [29] and Loehr [42] show, using slightly different language, that the assignment of the diagonal and dinv of every car ($\mathrm{area}(c, \pi)$ and $\mathrm{dinv}(c, \pi)$, respectively, for every car c) in this way uniquely defines a parking function. In particular, if for a particular parking function π, $a_c = \mathrm{area}(c, \pi)$ and $b_c = \mathrm{dinv}(c, \pi)$, then only π has area sequence $(a_1, a_2, \cdots, a_n)$ and dinv sequence $(b_1, b_2, \cdots, b_n)$. Similarly, they give a recursion on parking functions using the area and pmaj statistic which shows that a parking function π' is uniquely defined by the sequences $(\mathrm{pmaj}(c, \pi'))_{c=1}^{n}$ and $(\mathrm{area_p}(c, \pi'))_{c=1}^{n}$. Moreover, they show that if π had area sequence $(a_1, a_2, \cdots, a_n)$ and dinv sequence $(b_1, b_2, \cdots, b_n)$, then there exists (uniquely by the previous comment) a parking function π' with pmaj sequence $(a_1, a_2, \cdots, a_n)$ and $\mathrm{area_p}$ sequence $(b_1, b_2, \cdots, b_n)$. We define $\Gamma(\pi) = \pi'$. Thus we have the theorem:

Theorem 6 (Loehr [42]).

$$\sum_{\pi \in PF_n} t^{\mathrm{area}(\pi)} q^{\mathrm{dinv}(\pi)} = \sum_{\pi \in PF_n} t^{\mathrm{pmaj}(\pi)} q^{\mathrm{area}(\pi)}. \tag{4}$$

Since the left-hand side of (4) is exactly $\langle \Pi_n(X; q, t), e_1^n \rangle$, we have the following corollary:

Corollary 1.

$$\langle \Pi_n(X; q, 1), e_1^n \rangle = \langle \Pi_n(X; 1, q), e_1^n \rangle.$$

The symmetry of the $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ makes it obvious that the Shuffle Theorem implies $\Pi_n(X; q, t) = \Pi_n(X; t, q)$, but the corollary above is the closest any direct combinatorial proof has come to proving the symmetry.

3.3.2 Affine Permutations, Hyperplanes, and an Alternate Description of Γ

While the above description of Γ was historically identified first and gives a great deal of information about the three historically studied statistics on the parking functions, there is another description of Γ as the composition of two bijections. These bijections were originally defined as bijections from parking functions to regions of the Shi hyperplane arrangement, but can be more explicitly defined in terms of a subset of affine permutations which are in bijection with these regions. Much of this work is first stated in Armstrong [1] and Armstrong and Rhoades [2]. We base this section on the more general bijections in Gorsky et al. [22], adding a few minor changes documented below, and will return to the more general bijections later.

Definition (affine symmetric group)**.** Recall that the affine symmetric group $\tilde{\mathfrak{S}}_n$ can be identified with bijections $w : \mathbb{Z} \to \mathbb{Z}$ subject to the restrictions:

(1) For all integers x, $w(x) = w(n + x)$.
(2) $\sum_{i=1}^{n} w(i) = \frac{n(n+1)}{2}$.

Multiplication in the group is then composition of functions.

The affine permutations are generally represented using window notation $[w(1), \cdots, w(n)]$. Affine permutations are generated by $s_0, \cdots, s_{n-1}$ where:

$$\begin{aligned}
s_0 &= [0, 2, 3, 4, \cdots, n-1, n+1]\\
s_1 &= [2, 1, 3, 4, \cdots, n-1, n]\\
s_2 &= [1, 3, 2, 4, \cdots, n-1, n]\\
&\vdots\\
s_{n-1} &= [0, 1, 2, 3, 4, \cdots, n, n-1].
\end{aligned}$$

$\tilde{\mathfrak{S}}_n$ can be equivalently defined by relations on these generators:

(1) $s_i^2 = 1$.
(2) $s_i s_j = s_j s_i$ for $i - j \neq \pm 1 \mod n$.

(3) $s_i s_j s_i = s_j s_i s_j$ for $i - j = \pm 1 \mod n$.

Definition (m-restricted)**.** Call $w \in \tilde{\mathfrak{S}}_n$ m-restricted if for all $i < j$, $w(i) - w(j) \neq m$. Denote the set of m restricted affine permutations ${}^{m}\tilde{\mathfrak{S}}_n$.

Then to define the bijection $\tilde{\mathcal{A}} : PF_n \leftrightarrow {}^{n+1}\tilde{\mathfrak{S}}_n$ (similar to a map of Athanasiadis and Linusson [5] but here given in the language of Gorsky et al. [22]), we start with an assignment of "rank" to each of the cars. (In fact, it's easiest to assign a rank to each square in the lattice diagram, although we only care about the rank of the cells containing cars.) We assign rank according to the following convention:

(1) The bottom left cell has rank 0.
(2) Every step northward corresponds to an increase in rank of $n + 1$.
(3) Every eastward step corresponds to a decrease in rank of n.

(Note that the choice of rank of the bottom left cell is arbitrary, and there are different conventions depending on the author.)

Example 12. By this convention, looking at the cars in the parking function on the left of Fig. 6, car 7 has a rank of 0. The empty cell above 7 has a rank of $0 + 10 = 10$, so the cell containing car 3 has a rank of $10 - 9 = 1$. The following chart gives the ranks for all the cars in the parking function:

car	1	2	3	4	5	6	7	8	9
rank	7	8	1	5	11	13	0	21	15

To create $\tilde{\mathcal{A}}(\pi)$:

- Create a list of the ranks of each car in π (ordered by the cars):

$$w = [\text{rank}(1), \text{rank}(2), \cdots, \text{rank}(n)].$$

- w as created will generally not be an affine permutation since $\sum_{i=1}^{n} w(i) \neq \frac{n(n+1)}{2}$. Rescale w (adding $\frac{1}{n}\left(\frac{n(n+1)}{2} - \sum_{i=1}^{n} w(i)\right)$ to each term in the window).

The result of the rescaling is $\tilde{\mathcal{A}}(\pi)$.

Example 13. $7 + 8 + 1 + 5 + 11 + 13 + 0 + 21 + 15 = 81$ is the sum of the ranks of the cars in the parking function on the left of Example 6. Since our final permutation w should sum to 45, we need to subtract $36/9 = 4$ from each rank to get the final permutation. Thus $\tilde{\mathcal{A}}(\pi) = [3, 4, -3, 1, 7, 9, -4, 17, 11]$.

Remark. For this algorithm to work, it is necessary that the rescaling factor to be an integer. To see this is indeed the case, note that every cell in a row of a parking function has the same rank modulo n, no two rows contain ranks of the same modulus, and finally that each car is in a different row of the parking function. All of this is guaranteed by the fact that n and $n + 1$ are coprime. Why do we only get $n + 1$ restricted affine permutations? Cars with ranks that are exactly $n + 1$ apart

correspond to cars that are in the same column, and thus if $w(i) - w(j) = n + 1$, car i must be in the cell directly above car j and thus (by the column increasing condition) must correspond to a larger car.

Next, we define a second (much simpler) bijection $\tilde{\mathcal{PS}} :^{n+1} \tilde{\mathfrak{S}}_n \leftrightarrow PF_n$, where the result is a parking function in one-line notation (or equivalently the ith term gives the column of the ith car). In particular, for $1 \le i \le n$, we can define the

$$\tilde{\mathcal{PS}}(w)_i := 1 + \#\{j > i | 0 < w(i) - w(j) < n + 1\}.$$

It is important for this definition that we keep track of bounded inversions involving indices j that are outside the window.

Example 14. Continuing the previous example, notice that

$$\tilde{\mathcal{PS}} \circ \tilde{\mathcal{A}}(\pi) = \tilde{\mathcal{PS}}([3, 4, -3, 1, 7, 9, -4, 17, 11])$$

is the parking function $(4, 4, 2, 2, 3, 3, 1, 8, 4)$, which corresponds exactly to the parking function on the right-hand side of Fig. 6. For example, the 5th term in the resulting parking function is 3. This corresponds to the inversions between the 5th term in w and the 12th and 16th. Note that this does not include an inversion between the 5th and the 7th, since $7 - (-4) = 11 > 9 + 1$.

Our map $\tilde{\mathcal{A}}$ is closely related to the map $\mathcal{A}$ in Gorsky et al. [22]. Letting inv gives the inverse map on affine permutations, we've given a combinatorially simpler version (from this perspective at least) by defining $\tilde{\mathcal{A}} = \text{inv} \circ \mathcal{A}^{-1}$. Similarly our following map $\tilde{\mathcal{PS}} = \mathcal{PS} \circ \text{inv}$ in the language of the same paper, a map which they briefly refer to as $\mathcal{SP}$. Gorsky et al. [22] state that $\mathcal{PS} \circ \mathcal{A}^{-1}$ is equal to ζ when restricted to the Dyck paths (a statement which is essentially due in this case to Armstrong [1]). It's not hard to see that $\mathcal{PS} \circ \mathcal{A}^{-1} = \tilde{\mathcal{PS}} \circ \tilde{\mathcal{A}}$ nor that if we look at the resulting map on all parking functions, the result is actually Γ.

Both $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{PS}}$ (at least in this classical case) can be closely identified with classical maps from alcoves in a certain hyperplane arrangement, called the Shi arrangement, to the parking functions.

Definition. Let $H_{i,j}^k$ gives the hyperplane defined by $x_i - x_j = k$ restricted to

$$V = \{x \in \mathbb{R}^n | x_1 + \cdots + x_n = 0\}.$$

Then, the affine braid arrangement is the set of hyperplanes

$$\tilde{B}_n = \{H_{i,j}^k : 0 < i < j \le n, k \in \mathbb{Z}\}.$$

The Shi arrangement is the set of hyperplanes

$$\mathcal{S}_n = \{H_{i,j}^k : 0 < i < j \le n, k \in \{0, 1\}\}.$$

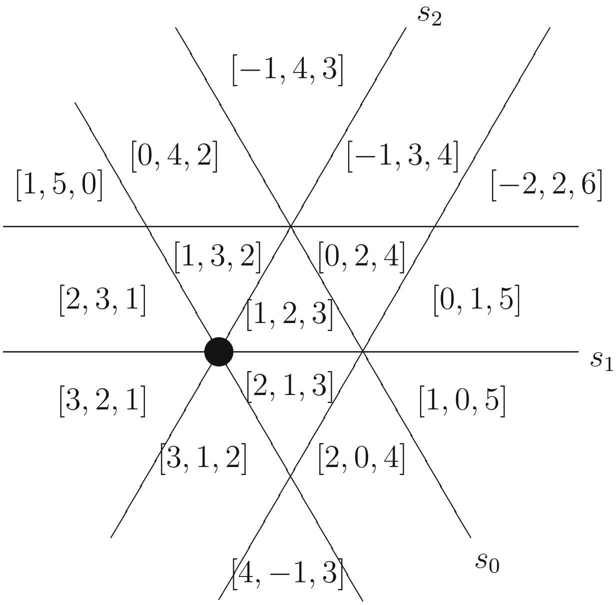


Fig. 7 Shi arrangement, with alcoves labeled by 4 restricted affine permutations of 3

Alcoves (connected regions bounded by the hyperplanes) in $\tilde{B}_n$ are in bijection with affine permutations of size n; alcoves in $\mathcal{S}_n$ are in bijection with only those permutations which are $n+1$ restricted. The "fundamental alcove" satisfying

$$x_1 > x_2 > \cdots > x_n > x_1 - 1$$

corresponds to the identity permutation $[1, 2, 3, \cdots, n]$. From there, new alcoves are labeled by reflection across $H^1_{1,n}$, corresponding to multiplication on the left by s_0, and $H^0_{i,i+1}$,corresponding to multiplication on the left by s_i. (Some sources, including Gorsky et al. [22], but not Armstrong [1] biject $\mathcal{S}_n$ to the $n+1$ bounded affine permutations, which are the inverse of the $n+1$ restricted permutations. In this context, one must label the regions by the same method, except that reflections across hyperplanes correspond to multiplication on the right, not the left.) If we then compose this labeling with $\tilde{\mathcal{A}}$, we get one labeling of the alcoves of the Shi arrangement by parking functions. This labeling is quite similar (but not identical to) the labeling of Athanasiadis and Linusson [5] of the Shi arrangement by parking functions. Similarly, we can label the regions by composing our labeling with $\tilde{\mathcal{PS}}^{-1}$. The result (in this case) is better known as the Pak-Stanley labeling Stanley [56]. The region where $x_1 > \cdots > x_n$ is labeled by increasing partitions corresponds under the Pak-Stanley labeling to Dyck paths.

Example 15. See Fig. 7 for a concrete example of the labeling.

4 Proof of the Shuffle Theorem

As is perhaps not surprising for a theorem two decades in the making, many researchers have played important roles in its proof, and we highlight a few of these results, including those results particularly important to the ultimate proof of the Shuffle Theorem.

A first key result, leading to the Shuffle Theorem, is the following result in Haiman [33]:

Theorem 7. *Let $DH_n[X; q, t]$ give the bigraded Frobenius characteristic of the diagonal harmonics. Then*

$$DH_n[X; q, t] = \nabla e_n.$$

The operator ∇ here is an eigenoperator of the Macdonald $\tilde{H}_\mu[X; q, t]$ first defined by François Bergeron. (See Bergeron et al. [9] for an extended discussion of the operator.) If μ' is the conjugate partition of μ, and

$$n(\mu) = \sum_{i=1}^{k} \mu_i (i - 1),$$

then ∇ is defined by

$$\nabla \tilde{H}_\mu[X; q, t] = t^{n(\mu)} q^{n(\mu')} \tilde{H}_\mu[X; q, t].$$

Haiman's proof of Theorem 7 is involved, requiring background both in algebraic geometry (relating to the Hilbert scheme of points in the plane) and algebraic combinatorics; parts of the proof were suggested earlier by Procesi and the formula was first conjectured in Garsia and Haiman [14].

Once Theorem 7 had been established, attention turned to establishing that

$$\nabla e_n = \Pi_n[X; q, t], \tag{5}$$

with the ultimate goal of showing, using Theorem 7, that the Frobenius characteristic of the diagonal harmonics was a weighted sum of parking functions. Progress toward the proof began to follow the same general steps:

(1) On the symmetric function side, several refinements of e_n were identified, including $e_n = \sum_k E_{n,k}$ in Garsia and Haglund [15], and $e_n = \sum_{\alpha \models n} C_\alpha$ in Haglund et al. [24], which after application of ∇, appeared to have nice Schur function expansions.
(2) Simultaneously, on the combinatorial side, a subset of the parking functions were identified which corresponded to the symmetric function refinement.
(3) Finally, an inner product with a product of elementary and complete homogeneous symmetric functions was taken with both the symmetric function and combinatorial refinement and the resulting polynomials were shown to be equal.

> On the combinatorial side, such an inner product just restricts the parking functions to those whose words were a certain type of shuffle.

Results of this sort were offered as evidence for the Shuffle Theorem, since they implied that some coefficients in the expansion of ∇e_n and $\Pi_n[X; q, t]$ in various symmetric function bases were the same, and the ultimate goal was to complete the process for an arbitrary length product of h's (or e's), thereby proving (5).

Besides the previously mention proof of the q, t Catalan, the first paper of the type was Haglund [26], who showed that

$$\langle \nabla e_n, e_a h_b \rangle = \left\langle \sum_{\pi \in PF_n} t^{\text{area}(\pi)} q^{\text{dinv}(\pi)} F_{\text{ides}(PF)}, e_a h_b \right\rangle$$

by refining the conjecture to identify those parking functions with k cars on the main diagonal.

The results were improved with a conjecture of Haglund et al. [24]:

$$\nabla C_\alpha = \sum_{\substack{\pi \in PF_n \\ touch(\pi)=\alpha}} t^{\text{area}(\pi)} q^{\text{dinv}(\pi)} F_{\text{ides}(PF)}. \tag{6}$$

Here C_α is a symmetric function created by the modified Hall–Littlewood vertex operators and $touch(\pi)$ is the "touch composition" of a parking function.

Definition (touch composition)**.** In particular, split the cars of π into an ordered set partition, adding cars by row and starting a new set every time the Dyck path returns to the main diagonal. The size of the sets gives a composition of n.

Example 16. In Fig. 2, the composition of the parking function on the left is $[1, 4, 2, 1, 1]$, counting the sets $(\{7\}, \{3, 5, 8, 6\}, \{4, 9\}, \{1\}, \{2\})$.

Using (6), recursions were found in Hicks [35], Garsia et al. [18], and Garsia et al. [19] which demonstrated the conjecture was true after taking various inner products, but there seemed to be no hope of extending this method of attack to prove the conjecture, without further refinement.

Recently, Carlsson and Mellit [12] announced a proof of the Shuffle Theorem. Their 35 page proof was expounded on in a 60 page, nearly self contained, series of lecture notes Haglund and Xin [25]. As the length of these both suggest, the proof is quite involved. Moreover, while the mathematics is a combination of symmetric function manipulation and bijective combinatorics, many of the details are quite technical. To the interested reader with background in algebraic combinatorics, we suggest Haglund and Xin [25] as a worthy resource. The proof requires a number of plethystic manipulations; beyond the lecture notes, the uninitiated reader may wish to consult Loehr and Remmel [43] for a better understanding of the useful notation. We make the following observations about the proof for the reader interested in a few of the details.

Carlsson and Mellit's proof picks up where Haiman's proof left off, in that they show:

Theorem 8 (Carlsson and Mellit [12])**.**

$$\nabla e_n = \Pi_n[X; q, t].$$

In fact, their proof refines and expands the conjecture of Haglund, Morse, and Zabrocki stated in (6) above. Carlsson and Mellit do not work directly with the right-hand side of (6) but rather with its image under the ζ map. Additionally, they use the opposite ordering of cars everywhere, so that columns are decreasing from bottom to top, the relative car size conditions on dinv are reversed as are the corner-based restrictions in DLD_n, etc. Since they work after applying the ζ map, they first must find an analogue of touch composition in the image of ζ; their definition of $touch'$ is (by necessity) considerably more complicated than $touch$. They call the image of the right-hand side of (6) under ζ, $D_\alpha(q,t)$, and their proof directly shows then that $D_\alpha(q,t) = \nabla C_\alpha$.

As with previous progress on the Shuffle Conjecture, a key idea in the proof is that further refining the conjecture makes it easier to prove. Carlsson and Mellit specifically identify symmetric function operators which give the weighted sum of all parking functions with a given Dyck path, further identifying even partial Dyck paths in some well-defined sense. In particular, they identify the set of DLD_n with a given path, and define "raising and lowering" operators (d_+ and d_-), which correspond in the symmetric functions to adding an east or a north step to a partial path. Expressing $D_\alpha(q,t)$ more cleanly in terms of these operators applied to a recursively defined polynomial (i.e., rather than as a sum of Dyck paths) leads to an expression that they show can be rewritten as ∇C_α; the mechanics of showing this equivalence is quite involved, involving a careful analysis of a number of operators and relations, and is not "just" a string of equalities.

5 Extensions of the Shuffle Theorem

Several natural extensions of the Shuffle Theorem are areas of active research. In this final section, we'll briefly describe several of these and give references for more complete details.

5.1 Rational Shuffle Theorem

One of the better known and more natural extensions of the Shuffle Theorem is the Rational Shuffle Theorem. From the perspective of algebraic combinatorics, the conjecture can be motivated from a natural extension of the Γ map as given in terms

of affine permutations and hyperplane arrangements. Gorsky et al. [22] show that $\tilde{\mathcal{PS}}$ and $\tilde{\mathcal{A}}$ can naturally be extended to the set of not just $n+1$ restricted affine permutations, but the set of m restricted permutations when m and n are relatively coprime. The resulting parking function image or preimage naturally becomes the set of so called rational parking functions, $PF_{m/n}$ which can similarly be pictured as occurring in $m \times n$ lattices containing paths of north and east steps which remain above the line $y = n/mx$. Again, one can determine the area of $\pi \in PF_{m/n}$ as the number of squares below the path and strictly above $y = (n/m)x$. Moreover, $\mathrm{dinv}(\pi)$ can be defined as $\mathrm{area}(\Gamma(\pi))$. A slight complication in this definition is that $\tilde{\mathcal{PS}}$, when it was defined, was not known to be bijective for all m, n, but recent work of Thomas and Williams [57] suggests this is at least the case when π is a parking function with reading word $[n, n-1, \cdots, 1]$. Moreover an equivalent (if far less elegant) definition for dinv can be worked out directly on π, as was first defined in Hikita [36] and was simplified in Hicks and Leven [34]. Together this suggests one could consider a combinatorial sum of weighted rational parking functions:

$$\Pi_{m,n}[X; q, t] = \sum_{\pi \in PF_{m/n}} t^{\mathrm{area}(\pi)} q^{\mathrm{dinv}(\pi)} F_{\mathrm{ides}(\pi)},$$

with some natural statistics generalizing those in the classical case. A reasonable question is whether one *should* study the sum (i.e., whether the sum is of independent interest). By expanding our attention from the Shuffle Theorem to this more general polynomial, a number of interesting additional connections become apparent. Hikita first investigated $\Pi_{m,n}[X; q, t]$ as the Frobenius image of a module arising naturally from the study of affine Springer fibers. Moreover, in Gorsky and Negut [20] a conjectured symmetric function is identified, leading to a natural analogue of the Shuffle Theorem:

Theorem 9 (Rational Shuffle Theorem)**.**

$$Q_{m,n}(-1)^n = \Pi_{m,n}[X; q, t].$$

A convincing reason to investigate $Q_{m,n}(-1)^n$ is its conjectured connections to irreducible modules in the rational Cherednik algebra as explained in Gorsky et al. [21] and its proven connections (see both Gorsky et al. [21] and Oblomkov et al. [46]) to knot invariants of the torus, where $Q_{m,n}(-1)^n$ gives a concrete way of computing the "superpolynomial knot invariant" (which generalizes the well-known HOMFLY polynomial) for torus knots. For the combinatorial audience, we suggest Haglund [23] for an overview of these connections, Bergeron et al. [10] and Bergeron et al. [11] for a clear statement of the conjecture, an explicit definition of $Q_{m,n}(-1)^n$, and what is known when m and n are not relatively coprime, and finally Mellit [45], for his recently released proof of the conjecture.

5.2 *Additional Conjectures and Theorems Related to* ∇

The Macdonald eigenoperator ∇ first appeared in the context of the Shuffle Theorem, applied naturally to e_n. Since then, experiments have shown that ∇ applied to many natural symmetric function basis elements is Schur positive; the result has been additional nice combinatorial formulas, currently either proven or conjectured. A nice summary of most of these conjectures occurs in Loehr and Warrington [41], although several nice works inspired by recent developments on the Shuffle Theorem occurred after its publication, including for example Sergel [53] and Wilson [58].

Additionally, it seems that ∇ itself has an interesting generalization, as first defined in Bergeron et al. [9]. Recall that ∇ is defined by

$$\nabla \tilde{H}_\mu[X; q, t] = t^{n(\mu)} q^{n(\mu')} \tilde{H}_\mu[X; q, t].$$

Next, if μ is a partition of k, $t^{n(\mu)} q^{n(\mu')}$ is equal to $e_k[B_\mu]$, where B_μ is a weighted sum of cells in the Ferrer's diagram of μ and $[\cdot]$ indicates plethystic substitution. Then several works have examined a new family of eigenoperators of the modified Macdonald polynomials also defined in Bergeron et al. [9] for a general symmetric function f by:

$$\Delta_f \tilde{H}_\mu[X; q, t] = f[B_\mu] \tilde{H}_\mu[X; q, t].$$

It follows that $\nabla f = \Delta_{e_k} f$ if f is a homogeneous symmetric function of degree k. Interestingly, for various natural choices of f, $\Delta_f e_n$ appears to still be Schur positive, with Haglund et al. [31] giving one of the best-known conjectures in this area for a combinatorial interpretation of the polynomial and Rhoades [49] and Romero [50] providing evidence for it. In some cases (such as Rhoades and Wilson [48]), there are conjectured modules whose bigraded Frobenius characteristics are believed to give a representation theoretic interpretation for the symmetric function as well. There is not yet one known combinatorial perspective for studying either ∇f or $\Delta_f e_n$ as f varies over the symmetric function bases.

Several additional generalizations of the diagonal harmonics can be found in Bergeron [6], which examines generalizing the $\mathfrak{S}_n$ action to a finite complex reflection group and working with an arbitrary number of sets of variables to define a new family of spaces generalizing the diagonal harmonics. The work shows the resulting spaces have Frobenius characteristics with some nice common properties, but does not give a general analogue of the Shuffle Theorem. Bergeron and Préville-Ratelle [8] specifically examine the case of three sets of variables with a simultaneous $\mathfrak{S}_n$ action and conjecture a natural analogue of the Shuffle Theorem, although without proposing a necessary third statistic. Since the closely related Hilbert scheme of points in the plane is no longer smooth for more than two sets of variables, Haiman [33] comments that the algebraic geometric techniques used to explore the diagonal harmonics do not readily generalize to the case of additional sets of variables.

Acknowledgements The author would like to express her gratitude for the many helpful remarks of the anonymous reviewer.

References

1. D. Armstrong, Hyperplane arrangements and diagonal harmonics, in *23rd International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2011)*, pp. 39–50. Discrete Mathematics and Theoretical Computer Science (2011)
2. D. Armstrong, B. Rhoades, The Shi arrangement and the Ish arrangement. Trans. Am. Math. Soc. **364**(3), 1509–1528 (2012)
3. D. Armstrong, N.A. Loehr, G.S. Warrington, Sweep maps: a continuous family of sorting algorithms. Adv. Math. **284**, 159–185 (2015)
4. S. Assaf, Toward the Schur expansion of Macdonald polynomials. Electron. J. Combin. **25**(2), Paper 2.44 (2018)
5. C.A. Athanasiadis, S. Linusson, A simple bijection for the regions of the Shi arrangement of hyperplanes. Discret. Math. **204**(1–3), 27–39 (1999)
6. F. Bergeron, Multivariate diagonal coinvariant spaces for complex reflection groups. Adv. Math. **239**, 97–108 (2013). ISSN 0001-8708
7. F. Bergeron, *Algebraic Combinatorics and Coinvariant Spaces* (CRC Press, Boca Raton, 2009)
8. F. Bergeron, L.-F. Préville-Ratelle, Higher trivariate diagonal harmonics via generalized Tamari posets. J. Comb. **3**(3), 317–341 (2012)
9. F. Bergeron, A. Garsia, M. Haiman, G. Tesler, Identities and positivity conjectures for some remarkable operators in the theory of symmetric functions. Methods Appl. Anal. **6**(3), 363–420 (1999)
10. F. Bergeron, A. Garsia, E.S. Leven, G. Xin, Compositional (km, kn)-shuffle conjectures. Int. Math. Res. Not. **2016**(14), 4229–4270 (2015)
11. F. Bergeron, A. Garsia, E.S. Leven, G. Xin, Some remarkable new plethystic operators in the theory of Macdonald polynomials. J. Comb. **7**(4), 671–714 (2016)
12. E. Carlsson, A. Mellit, A proof of the shuffle conjecture. J. Am. Math. Soc. **31**(3), 661–697 (2018)
13. D. Foata, J. Riordan, Mappings of acyclic and parking functions. Aequ. Math. **10**(1), 10–22 (1974)
14. A.M. Garsia, M. Haiman, Some natural bigraded S_n-modules and q, t-Kostka coefficients. Electron. J. Combin. **3** (1996)
15. A. Garsia, J. Haglund, A positivity result in the theory of Macdonald polynomials. Proc. Natl. Acad. Sci. **98**(8), 4313–4316 (2001)
16. A.M. Garsia, J. Haglund, A proof of the q, t-Catalan positivity conjecture. Discret. Math. **256**(3), 677–717 (2002)
17. A.M. Garsia, M. Haiman, A graded representation model for Macdonald's polynomials. Proc. Natl. Acad. Sci. **90**(8), 3607–3610 (1993)
18. A.M. Garsia, G. Xin, M. Zabrocki, Hall-Littlewood operators in the theory of parking functions and diagonal harmonics. Int. Math. Res. Not. **2012**(6), 1264–1299 (2011)
19. A.M. Garsia, G. Xin, M. Zabrocki, A three shuffle case of the compositional parking function conjecture. J. Comb. Theory Ser. A **123**(1), 202–238 (2014)
20. E. Gorsky, A. Negut, Refined knot invariants and Hilbert schemes. J. Math. Pures Appl. (9) **104**(3), 403–435 (2015)
21. E. Gorsky, A. Oblomkov, J. Rasmussen, V. Shende, Torus knots and the rational DAHA. Duke Math. J. **163**(14), 2709–2794 (2014)
22. E. Gorsky, M. Mazin, M. Vazirani, Affine permutations and rational slope parking functions. Trans. Am. Math. Soc. **368**(12), 8403–8445 (2016)
23. J. Haglund, The combinatorics of knot invariants arising from the study of Macdonald polynomials. *Recent Trends in Combinatorics* (Springer, Berlin, 2016), pp. 579–600
24. J. Haglund, J. Morse, M. Zabrocki, A compositional shuffle conjecture specifying touch points of the Dyck path. Canad. J. Math. **64**(4), 822–844 (2012). ISSN 0008-414X
25. J. Haglund, G. Xin, Lecture notes on the Carlsson-Mellit proof of the shuffle conjecture (2017), arXiv:1705.11064

26. J. Haglund, A proof of the q, t-Schröder conjecture. Int. Math. Res. Not. **2004**(11), 525–560 (2004)
27. J. Haglund, The genesis of the Macdonald polynomial statistics. Séminaire Lotharingien de Combinatoire **54**, B54Ao (2006)
28. J. Haglund, *The q, t-Catalan Numbers and the Space of Diagonal Harmonics*, vol. 41 (American Mathematical Society, Providence, 2008)
29. J. Haglund, N. Loehr, A conjectured combinatorial formula for the Hilbert series for diagonal harmonics. Discret. Math. **298**(1), 189–204 (2005)
30. J. Haglund, M. Haiman, N. Loehr, A combinatorial formula for Macdonald polynomials. J. Am. Math. Soc. **18**(3), 735–761 (2005)
31. J. Haglund, J. Remmel, A.T. Wilson, The delta conjecture. Trans. Am. Math. Soc. **370**(6), 4029–4057 (2018)
32. M. Haiman, Hilbert schemes, polygraphs and the Macdonald positivity conjecture. J. Am. Math. Soc. **14**(4), 941–1006 (2001)
33. M. Haiman, Vanishing theorems and character formulas for the Hilbert scheme of points in the plane. Inventiones Math. **149**(2), 371–407 (2002)
34. A. Hicks, E. Leven, A simpler formula for the number of diagonal inversions of an (m, n)-parking function and a returning fermionic formula. Discrete Math. **338**(3), 48–65 (2015)
35. A.S. Hicks, Two parking function bijections: a sharpening of the q, t-Catalan and Shröder theorems. Int. Math. Res. Not. **2012**(13), 3064–3088 (2011)
36. T. Hikita, Affine Springer fibers of type A and combinatorics of diagonal coinvariants. Adv. Math. **263**, 88–122 (2014)
37. D.E. Knuth, Linear probing and graphs. Algorithmica **22**(4), 561–568 (1998)
38. A.G. Konheim, B. Weiss, An occupancy discipline and applications. SIAM J. Appl. Math. **14**(6), 1266–1274 (1966)
39. A. Lascoux, B. Leclerc, J.-Y. Thibon, Ribbon tableaux, Hall-Littlewood functions, quantum affine algebras, and unipotent varieties. J. Math. Phys. **38**(2), 1041–1068 (1997)
40. B. Leclerc, J.-Y. Thibon, Littlewood-Richardson coefficients and Kazhdan-Lusztig polynomials. *Combinatorial Methods in Representation Theory*, Advanced Studies in Pure Mathematics, vol. 28 (Citeseer, 1998)
41. N.A. Loehr, G.S. Warrington, Nested quantum Dyck paths and $\nabla(s_\lambda)$. Int. Math. Res. Not. **5** (2008)
42. N.A. Loehr, Combinatorics of q, t-parking functions. Adv. Appl. Math. **34**(2), 408–425 (2005)
43. N.A. Loehr, J.B. Remmel, A computational and combinatorial exposé of plethystic calculus. J. Algebr. Comb. **33**(2), 163–198 (2011)
44. I.G. Macdonald, *Symmetric Functions and Hall Polynomials* (Oxford University Press, Oxford, 1998)
45. A. Mellit, Toric braids and (m, n)-parking functions (2016), arXiv:1604.07456
46. A. Oblomkov, V. Shende, The Hilbert scheme of a plane curve singularity and the homfly polynomial of its link. Duke Math. J. **161**(7), 1277–1303 (2012)
47. R. Pyke, The supremum and infimum of the poisson process. Ann. Math. Stat. **30**(2), 568–576 (1959)
48. B. Rhoades, A.T. Wilson, Tail positive words and generalized coinvariant algebras. Electron. J. Combin. **24**(3), Paper 3.21, 29 (2017)
49. B. Rhoades, Ordered set partition statistics and the delta conjecture. J. Comb. Theory Ser. A **154**, 172–217 (2018)
50. M. Romero, The delta conjecture at $q = 1$. Trans. Am. Math. Soc. **369**(10), 7509–7530 (2017)
51. B.E. Sagan, *The Symmetric Group*. Volume 203 of Graduate Texts in Mathematics (Springer, New York, 2001), 2nd edn. Representations, combinatorial algorithms, and symmetric functions
52. A. Schilling, M. Shimozono, D. White, Branching formula for q-Littlewood-Richardson coefficients. Adv. Appl. Math. **30**(1–2), 258–272 (2003)
53. E. Sergel, A proof of the square paths conjecture. J. Comb. Theory Ser. A **152**, 363–379 (2017)

54. N.J.A. Sloane, The on-line encyclopedia of integer sequences, http://oeis.org. Sequence A000272
55. R.P. Stanley, *Enumerative Combinatorics. Vol. 2*, Cambridge Studies in Advanced Mathematics, vol. 62 (Cambridge University Press, Cambridge, 1999)
56. R.P. Stanley, Hyperplane arrangements, interval orders, and trees. Proc. Natl. Acad. Sci. **93**(6), 2620–2625 (1996)
57. H. Thomas, N. Williams, Sweeping up zeta. Sel. Math. (2018)
58. A.T. Wilson, Torus link homology and the nabla operator. J. Comb. Theory Ser. A **154**, 129–144 (2018)
59. C.H. Yan, Parking functions, *Handbook of Enumerative Combinatorics*, Discrete Mathematics and Applications (CRC Press, Boca Raton, FL, 2015), pp. 835–893

On Positivity of Ehrhart Polynomials

Fu Liu

Abstract Ehrhart discovered that the function that counts the number of lattice points in dilations of an integral polytope is a polynomial. We call the coefficients of this polynomial Ehrhart coefficients and say a polytope is Ehrhart positive if all Ehrhart coefficients are positive (which is not true for all integral polytopes). The main purpose of this chapter is to survey interesting families of polytopes that are known to be Ehrhart positive and discuss the reasons from which their Ehrhart positivity follows. We also include examples of polytopes that have negative Ehrhart coefficients and polytopes that are conjectured to be Ehrhart positive and pose a few relevant questions.

1 Introduction

A *polyhedron* in the D-dimensional Euclidean space $\mathbb{R}^D$ is the solution set of a finite set of linear inequalities:

$$P = \left\{ \boldsymbol{x} \in \mathbb{R}^D \; : \; \sum_{j=1}^{D} a_{i,j} x_j \leq b_i \text{ for } i \in I \right\},$$

where $a_{i,j} \in \mathbb{R}$, $b_i \in \mathbb{R}$, and I is a finite set of indices. A *polytope* is a bounded polyhedron. Equivalently, a *polytope* in $\mathbb{R}^D$ can also be defined as the convex hull of finitely many points in $\mathbb{R}^D$. We assume readers are familiar with basic definitions such as *faces* and *dimensions* of polytopes as presented in [108]. In this paper, the letter d usually denotes the dimension of a polytope and D denotes the dimension of the ambient space. For majority of the examples presented here, we either have $d = D$ or $d = D - 1$.

F. Liu (✉)
Department of Mathematics, University of California, One Shields Avenue,
Davis, CA 95616, USA
e-mail: fuliu@math.ucdavis.edu

H. Barcelo et al. (eds.), *Recent Trends in Algebraic Combinatorics*,
Association for Women in Mathematics Series 16,
https://doi.org/10.1007/978-3-030-05141-9_6

A *lattice point* or an *integral point* is a point in $\mathbb{Z}^D$. Counting lattice points inside polytopes is a fundamental and useful step in many mathematical analyses. A lot of combinatorial structures can be counted as lattice points of polytopes. For example, matchings on graphs [63], t-designs [74], (semi-)magic squares [10, Chapter 6], and linear extensions of posets [102] are all of this form. Counting lattice points not only appears in combinatorial problems, it also appears, for instance, in the context of representation theory [52, 89], algebraic geometry [37], statistics [34, 36], and number theory [7, 75].

One approach to study the question of computing the number of lattice points in a polytope P is to consider a more general counting problem: For any nonnegative integer t, let $tP := \{t\boldsymbol{x} \ : \ \boldsymbol{x} \in P\}$ be the tth *dilation of* P, and then consider the function

$$i(P, t) := |tP \cap \mathbb{Z}^D|,$$

which counts the number of lattice points in tP. We say two polytopes P and Q are *unimodularly equivalent*[1] if there exists an affine transformation from the affine hull $\mathrm{aff}(P)$ of P to the affine hull $\mathrm{aff}(Q)$ of Q that induces a bijection from lattice points in $\mathrm{aff}(P)$ to lattice points in $\mathrm{aff}(Q)$. Such an affine transformation is called a *unimodular transformation*. It is easy to see that if two polytopes P and Q are unimodularly equivalent, then $i(P, t) = i(Q, t)$.

An *integral polytope* (or a *lattice polytope*) is a polytope whose vertices are lattice points. In the 1960s, Eugène Ehrhart [35] discovered that the function $i(P, t)$ has nice properties when P is an integral polytope.

Theorem 1.1 (Ehrhart)**.** *For any integral d-polytope P, the function $i(P, t)$ is always a polynomial (with real coefficients) of degree d in t.*

Thus, we call $i(P, t)$ the *Ehrhart polynomial* of an integral polytope P, and the coefficients of $i(P, t)$ the *Ehrhart coefficients* of P. Note that Ehrhart's theorem can be extended to rational polytopes with the concept of a quasi-polynomial; however, we will focus on integral polytopes in this chapter.

There is much work on the Ehrhart coefficients of integral polytopes. In the 1990s, many people studied the problem of counting lattice points inside integral (or more generally rational) polytopes [17, 22, 50, 80, 85] by using the theory of toric varieties. Although explicit formulas for coefficients of Ehrhart polynomials can be deduced from these results, they are often quite complicated. Only three coefficients of $i(P, t)$ have simple known forms for arbitrary integral polytopes P: The leading coefficient is equal to the normalized volume of P, the second coefficient is one half of the sum of the normalized volumes of facets, and the constant term is always 1.

Although these three coefficients can be described in terms of volumes (recall 1 is the normalized volume of a point) and thus are positive, it is not true that all the coefficients of $i(P, t)$ are positive. (The first counterexample comes up in dimension 3 known as the *Reeve tetrahedron*; see Sect. 4.1.) We say a polytope has *Ehrhart*

[1] Unimodular equivalence is sometimes called *integral equivalence*, e.g., in [79].

positivity or is *Ehrhart positive* if it has positive Ehrhart coefficients. It is natural to ask the following question:

Question 1.2. Which families of integral polytopes have Ehrhart positivity?

This turns out to be a challenging question. Even though multiple families of polytopes have been shown to be Ehrhart positive in the literature, the techniques involved are (almost) all different. In Sect. 2, we will survey families of polytopes with the Ehrhart positivity property, discussing different reasons why they have this property. In particular, as a consequence of the techniques discussed in Sect. 2.4, one can show that any combinatorial type of rational polytopes can be realized as an integral polytope that is Ehrhart positive (see Theorem 2.4.8). This result indicates that Ehrhart positivity is *not* a combinatorial property. Therefore, it is desirable to find more geometric methods to prove Ehrhart positivity. In Sect. 3, we introduce such a tool called *McMullen's formula*, which we use to give a refinement of Ehrhart positivity, called α-positivity. We then use this tool to attack the Ehrhart positivity conjecture of "generalized permutohedra," a family of polytopes introduced by Postnikov [82], and report partial progress on this conjecture. In Sect. 4, we include negative results on Question 1.2, presenting examples with negative Ehrhart coefficients. In particular, we will discuss progress on a question asked and studied by Hibi, Higashitani, Tsuchiya, and Yoshida [47] on all possible sign patterns of Ehrhart coefficients (see Sect. 4.2). Note that this question can be considered to be a refinement of Question 1.2. Finally, in Sect. 5, we include various conjectures on Ehrhart positivity and pose related questions.

We finish our introduction with the following remark on the coefficients of the h^*-*polynomial*, which is closely related to the Ehrhart polynomials.

Remark on h^*-Vector. One method of proving Ehrhart's theorem (Theorem 1.1) is by considering the *Ehrhart series* of a d-dimensional integral polytope P:

$$\mathrm{Ehr}_P(z) := \sum_{t\geq 0} i(P,t)z^t.$$

It turns out that Theorem 1.1 is equivalent to the existence of a polynomial $h^*_P(z)$ of degree at most d such that $h^*_P(1) \neq 0$ and

$$\mathrm{Ehr}_P(z) = \frac{h^*_P(z)}{(1-z)^{d+1}}.$$

See [94, Chapter 4] for a statement for the above equivalence result and a proof for Ehrhart's theorem.

We call $h^*_P(z)$ the h^*-*polynomial* of P, and the vector $(h^*_0, h^*_1, \ldots, h^*_d)$, where h^*_i is the coefficient of z^i in $h^*_P(z)$, the h^*-*vector* of P. One can recover the Ehrhart polynomial of a d-dimensional integral polytope P easily from its h^*-vector:

$$i(P,t) = \sum_{j=0}^{d} h_j^* \binom{t+d-j}{d}. \tag{1.1}$$

It is a well-known result due to Stanley that the entries in h^*-vectors are all nonnegative integers [93] in contrast to the fact that Ehrhart coefficients could be negative. As a consequence, positivity is not such an interesting question for h^*-polynomials. Nevertheless, active research has been conducted in other directions.

The most natural question probably is: For each d, can we give a complete characterization for all possible h^*-vector of d-dimensional integral polytopes? For $d = 2$, the answer was first provided in 1976 by Scott [90] known as *Scott's condition*. However, for $d \geq 3$, the question is wide open. A lot of work has been done in the literature on searching for inequalities and equalities satisfied by h^*-vectors. Most of them were discovered by Hibi [43–46] and Stanley [93, 99] in the 1990s using commutative algebra and combinatorial methods. In 2009, Stapledon [103] contributes more inequalities using the idea of degree and codegree of a polytope. Known equalities on h^*-vectors include

$$\sum_{i=0}^{d} h_i^* = d!\text{Vol}(P), \quad h_0^* = 1, \quad h_1^* = i(P) - (d+1), \quad h_d^* = |\text{relint}(P) \cap \mathbb{Z}^D|. \tag{1.2}$$

Please see [9, 103] for lists of known inequalities. Recently, instead of focusing on inequalities satisfied by all polytopes, much work has been done on finding inequalities for polytopes under certain constraints. For example, Treutlein [105] shows that the necessary statement of Scott's condition holds for any integral polytope whose h^*-polynomial is of degree at most 2, i.e., $h_i^* = 0$ for all $i > 2$. Most recently, Balletti and Higashitani [5] improve the result further to any integral polytope whose h^*-polynomial satisfies $h_3^* = 0$.

Another question that comes up a lot in the context of h^*-vector is the unimodality question. A sequence of real numbers $c_0, c_1, \ldots, c_d$ is *unimodal* if there exists $0 \leq j \leq d$ such that $c_0 \leq c_1 \leq \cdots \leq c_j \geq c_{j+1} \geq \cdots \geq c_d$. It is well known that a nonnegative sequence is unimodal if it has "no internal zeros" and is "log-concave," and furthermore, log-concavity follows from another property called "real-rootedness." Please see surveys by Stanley [98] and Brenti [16] on log-concave and unimodal sequences, and a survey by Brändén [13] with a more general discussion on unimodality, log-concavity, and real-rootedness. Recently, Braun [14] wrote a survey on unimodality problem of the h^*-vector of integral polytopes, discussing a wide range of tools (including but not limited to the techniques mentioned in the aforementioned surveys) to attack this problem. Finally, we would like to remark that even though Ehrhart coefficients and h^*-vectors are related by (1.1), there is no general implication between Ehrhart positivity and h^*-unimodality [60].

2 Polytopes with Ehrhart Positivity

In the literature, there are multiple interesting families of polytopes shown to be Ehrhart positive using very different techniques. In this section, we put together a collection of such families, separating them into four categories based on the reasons why they are Ehrhart positive. However, we make no attempt to give a comprehensive account of *all* families with this property. We also note that as the leading coefficient of $i(P, t)$ is the volume of P, one can often extract a formula for volume from descriptions for Ehrhart polynomials we give below. However, we will focus only on results on Ehrhart polynomials here and omit related formulas for volumes.

In this chapter, we use bold letters to denote both vectors and points in $\mathbb{R}^D$. For example, $\boldsymbol{e}_i$ denotes both the ith vector in the standard basis and the point $(0, \dots, 0, 1, 0, \dots, 0)$ where 1 is in the ith position.

For convenience, we use $\mathbb{N}$ to denote the set of nonnegative integers, and $\mathbb{P}$ the set of positive integers.

2.1 Products of Positive Linear Polynomials

In this part, we present families of polytopes whose Ehrhart polynomials can be described explicitly, which can be shown to have positive coefficients using the following naive lemma.

Lemma 2.1.1. *Suppose a polynomial $f(t)$ is either*

(a) a product of linear polynomials with positive coefficients, or
(b) a sum of products of linear polynomials with positive coefficients.

Then $f(t)$ has positive coefficients.

We start with the two simplest families of polytopes: *unit cubes* and *standard simplices*, whose Ehrhart polynomials fit into situation (a) of Lemma 2.1.1, and thus Ehrhart positivity follows. As these are the first examples of Ehrhart polynomials in this chapter, and the computations are straightforward, we include all the details. For most of the remaining examples we discuss in this paper, we only state the results without providing detailed proofs.

2.1.1 Unit Cubes

The d-dimensional *unit cube*, denoted by $\square_d$, is the convex hull of all points in $\mathbb{R}^d$ with coordinates in $\{0, 1\}$, i.e.,

$$\square_d := \operatorname{conv}\{\boldsymbol{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d \ : \ x_i = 0 \text{ or } 1 \text{ for } i = 1, 2, \dots, d\}.$$

It is easy to verify that the unit cube is the solution set to the following linear system of inequalities:

$$\square_d = \{\boldsymbol{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d \ : \ 0 \le x_i \le 1 \text{ for } i = 1, 2, \dots, d\},$$

Then for any $t \in \mathbb{N}$,

$$t\square_d = \{\boldsymbol{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d \ : \ 0 \le x_i \le t \text{ for } i = 1, 2, \dots, d\}.$$

Thus,

$$t\square_d \cap \mathbb{Z}^d = \{\boldsymbol{x} = (x_1, x_2, \dots, x_d) \in \mathbb{Z}^d \ : \ 0 \le x_i \le t \text{ for } i = 1, 2, \dots, d\}.$$

For each i, the number of integers x_i such that $0 \le x_i \le t$ is $t + 1$. Thus,

$$i(\square_d, t) = (t+1)^d.$$

2.1.2 Standard Simplices

The d-dimensional *standard simplex*, denoted by Δ_d, is the convex hull of all the elements in the standard basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_{d+1}$ of $\mathbb{R}^{d+1}$:

$$\Delta_d := \mathrm{conv}\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_{d+1}\}.$$

One checks that Δ_d can also be defined by the following linear system:

$$\sum_{j=1}^{d+1} x_j = 1, \text{ and } x_i \ge 0 \text{ for } i = 1, 2, \dots, d+1.$$

Hence, for any $t \in \mathbb{N}$,

$$t\Delta_d = \left\{ \boldsymbol{x} = (x_1, x_2, \dots, x_{d+1}) \in \mathbb{R}^{d+1} \ : \ \sum_{j=1}^{d+1} x_j = t, \text{ and } x_i \ge 0 \text{ for } i = 1, 2, \dots, d+1 \right\},$$

and

$$t\Delta_d \cap \mathbb{Z}^{d+1} = \left\{ \boldsymbol{x} = (x_1, x_2, \dots, x_{d+1}) \in \mathbb{Z}^{d+1} \ : \ \sum_{j=1}^{d+1} x_j = t, \text{ and } x_i \ge 0 \text{ for } i = 1, 2, \dots, d+1 \right\}.$$

Hence, $i(\Delta_d, t)$ counts the number of nonnegative integer solutions to

$$x_1 + x_2 + \cdots + x_{d+1} = t.$$

This is a classic combinatorial problem which is the same as counting the number of weak compositions of t into $d+1$ parts (see [94, Page 18]), and the answer is given by

$$i(\Delta_d, t) = \binom{t+d}{d}.$$

As we mentioned above, Ehrhart positivity of unit cubes and standard simplices follows from situation (a) of Lemma 2.1.1. Next, we present two families of examples with Ehrhart polynomials in the form of situation (b) of Lemma 2.1.1.

2.1.3 Pitman-Stanley Polytopes

Let $\boldsymbol{a} = (a_1, \ldots, a_d) \in \mathbb{N}^d$. The following polytope is introduced and studied by Pitman and Stanley [79]:

$$\mathcal{PS}_d(\boldsymbol{a}) := \left\{ \boldsymbol{x} \in \mathbb{R}^d \ : \ x_i \geq 0 \text{ and } \sum_{j=1}^{i} x_j \leq \sum_{j=1}^{i} a_i, \ \text{ for } i = 1, 2, \ldots, d \right\},$$

and hence we call it a *Pitman-Stanley polytope*.

Pitman and Stanley gave an explicit formula [79, Formula (33)] for computing the number of lattice points in $\mathcal{PS}_d(\boldsymbol{a})$, from which a formula for the Ehrhart polynomial of $\mathcal{PS}_d(\boldsymbol{a})$ immediately follows. Recall $\left(\!\binom{x}{y}\!\right) = \binom{x+y-1}{y}$.

Theorem 2.1.2 (Pitman-Stanley). *Let*

$$I_d := \left\{ \boldsymbol{i} = (i_1, i_2, \ldots, i_d) \in \mathbb{N}^d \ : \ \sum_{j=1}^{d} i_j = d, \ \textit{and} \ \sum_{j=1}^{k} i_j \geq k \textit{ for } k = 1, 2, \ldots d-1 \right\}.$$

Then the Ehrhart polynomial of $\mathcal{PS}_d(\boldsymbol{a})$ is given by

$$i(\mathcal{PS}_d(\boldsymbol{a}), t) = \sum_{\boldsymbol{i} \in I_d} \left(\!\binom{a_1 t + 1}{i_1}\!\right) \prod_{k=2}^{d} \left(\!\binom{a_k t}{i_k}\!\right). \tag{2.1}$$

For each $\boldsymbol{i}$, both $\left(\!\binom{a_1 t+1}{i_1}\!\right)$ and $\left(\!\binom{a_k t}{i_k}\!\right)$ are products of linear polynomials in t with positive coefficients, so it follows from Lemma 2.1.1/(b) that any Pitman-Stanley polytope $\mathcal{PS}_d(\boldsymbol{a})$ is Ehrhart positive.

Pitman-Stanley polytopes are contained in two different more general families of polytopes: flow polytopes and generalized permutohedra. For each of these two bigger families of polytopes, formulas for Ehrhart polynomials of some subfamily have been derived, generalizing Formula (2.1). We present results on flow polytopes in the next part below, while the results on generalized permutohedra are postponed

to Sect. 3.1.2 as part of a general discussion on the Ehrhart positivity conjecture of generalized permutohedra in Sect. 3.

2.1.4 Subfamilies of Flow Polytopes

Let G be a (loopless) directed acyclic connected graph on $[n+1] = \{1, 2, \dots, n+1\}$ such that each edge $\{i, j\}$ with $i < j$ is always directed from i to j. Hence, we denote the edge by (i, j) to indicate the orientation. For any $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in \mathbb{N}^n$, we associate to it another vector

$$\bar{\boldsymbol{a}} := \left(a_1, \dots, a_n, -\sum_{i=1}^n a_i\right). \tag{2.2}$$

An $\bar{\boldsymbol{a}}$-*flow* on G is a vector $\boldsymbol{f} = (f(e))_{e\in E(G)} \in \left(\mathbb{R}_{\geq 0}\right)^{E(G)}$ such that for $i = 1, 2, \dots, n$, we have

$$\sum_{e=(g,i)\in E(G)} f(e) + a_i = \sum_{e=(i,j)\in E(G)} f(e),$$

that is, the *netflow* at vertex i is a_i. Note these conditions imply that the netflow at vertex $n+1$ is $-\sum_{i=1}^n a_i$. The *flow polytope* $\mathcal{F}_G(\bar{\boldsymbol{a}})$ associated to G and the integer netflow $\bar{\boldsymbol{a}}$ is the set of all $\bar{\boldsymbol{a}}$-flows $\boldsymbol{f}$ on G.

Example 2.1.3. Let $G_d^{\mathcal{PS}}$ be the graph on $[d+1]$ with edge set

$$\{(i, i+1), (i, d+1) \ : \ i = 1, 2, \dots, d\}.$$

Baldoni and Vergne [3, Example 16] show that $\mathcal{F}_{\mathbb{G}_d^{\mathcal{PS}}}(\bar{\boldsymbol{a}})$ is unimodularly equivalent to the Pitman-Stanley polytope $\mathcal{PS}_d(\boldsymbol{a})$.

For each edge $e = (i, j)$ of G, we associate to it the positive type A_n root $\alpha(e) = \alpha(i, j) = \boldsymbol{e}_i - \boldsymbol{e}_j$. For any $\boldsymbol{b} \in \mathbb{Z}^{n+1}$, the *Kostant partition function* $\mathcal{KP}_G$ evaluated at $\boldsymbol{b}$ is

$$\mathcal{KP}_G(\boldsymbol{b}) := \#\left\{\boldsymbol{f} = (f(e))_{e\in E(G)} \in \mathbb{N}^{E(G)} \ : \ \sum_{e\in E(G)} f(e)\alpha(e) = \boldsymbol{b}\right\}.$$

It is straightforward to verify that for $\boldsymbol{a} \in \mathbb{N}^n$,

$$\mathcal{KP}_G(\bar{\boldsymbol{a}}) = |\mathcal{F}_G(\bar{\boldsymbol{a}}) \cap \mathbb{Z}^{E(G)}|,$$

i.e., $\mathcal{KP}_G(\bar{\boldsymbol{a}})$ counts the number of lattice points in the flow polytope $\mathcal{F}_G(\bar{\boldsymbol{a}})$. In the literature, various groups of people [3, 55, 70, 84] obtained formulas for Kostant

partition functions or equivalently the number of lattice points in flow polytopes. As a consequence, we can easily obtain formulas for the Ehrhart polynomial of $\mathcal{F}_G(\bar{\boldsymbol{a}})$.

Theorem 2.1.4 (Lidskii, Postnikov-Stanley, Baldoni-Vergne, Mészáros-Morales)**.** *Suppose G is a connected graph on the vertex set $[n+1]$, with m edges directed $i \to j$ if $i < j$, and with at least one outgoing edge at vertex i for $i = 1, \dots, n$. Let* out_k *(and* in_k*, respectively) denote the outdegree (and the indegree, respectively) of vertex k in G minus* 1.

Then for any $\boldsymbol{a} = (a_1, \dots, a_n) \in \mathbb{N}^n$, the Ehrhart polynomial of $\mathcal{F}_G(\bar{\boldsymbol{a}})$ is given by

$$
\begin{aligned}
& i(\mathcal{F}_G(\bar{\boldsymbol{a}}), t) \\
&= \sum_{\boldsymbol{j}} \prod_{k=1}^{n} \binom{a_k t + \mathrm{out}_k}{j_k} \cdot \mathcal{KP}_G(j_1 - \mathrm{out}_1, j_2 - \mathrm{out}_2, \dots, j_n - \mathrm{out}_n, 0), \qquad (2.3) \\
&= \sum_{\boldsymbol{j}} \prod_{k=1}^{n} \left(\!\binom{a_k t - \mathrm{in}_k}{j_k}\!\right) \cdot \mathcal{KP}_G(j_1 - \mathrm{out}_1, j_2 - \mathrm{out}_2, \dots, j_n - \mathrm{out}_n, 0), \qquad (2.4)
\end{aligned}
$$

where each summation is over all weak compositions $\boldsymbol{j} = (j_1, \dots, j_n)$ of $m - n$ that are $\geq (\mathrm{out}_1, \dots, \mathrm{out}_n)$ in dominance order.

We remark that Lidskii [55] gives a formula for computing Kostant partition functions associated to the complete graph K_{n+1}, which yields Formula (2.3) above with $G = K_{n+1}$. Postnikov and Stanley [84, unpublished] were the first to discover Formula (2.3) for arbitrary graphs G using the Elliott-MacMahon algorithm. Baldoni and Vergne [3] give a proof for both formulas in Theorem 2.1.4 using residue computation. Most recently, Mészáros and Morales [70] recover Baldoni-Vergne's result by extending ideas of Postnikov and Stanley on the Elliott-MacMahon algorithm and polytopal subdivisions of flow polytopes.

Formula (2.4) is useful in obtaining positivity results since

$$
\left(\!\binom{a_k t - \mathrm{in}_k}{j_k}\!\right) = \binom{a_k t - \mathrm{in}_k + j_k - 1}{j_k}
$$

is a product of linear polynomials in t with positive coefficients as long as $\mathrm{in}_k = 0$ or -1. Also, note that $\mathcal{KP}_G(j_1 - \mathrm{out}_1, j_2 - \mathrm{out}_2, \dots, j_n - \mathrm{out}_n, 0)$ is nonnegative and $i(\mathcal{F}_G(\bar{\boldsymbol{a}}), t) \neq 0$. The following result immediately follows from Lemma 2.1.1/(b).

Corollary 2.1.5. *Assume the hypotheses of Theorem 2.1.4. Assume further that for each vertex $i \in [n] = \{1, 2, \dots, n\}$, the indegree of i is either 0 or 1. Then the flow polytope $\mathcal{F}_G(\bar{\boldsymbol{a}})$ is Ehrhart positive.*

We remark that the graph $G_d^{\mathcal{PS}}$ defined in Example 2.1.3 satisfies the hypothesis of the above corollary. Hence, Ehrhart positivity of the Pitman-Stanley polytope is a special case of Corollary 2.1.5.

2.2 Roots with Negative Real Parts

In this part, we show examples with Ehrhart positivity using the following lemma. We use $\mathfrak{Re}(z)$ to denote the real part of a complex number z.

Lemma 2.2.1. *Let $p(t)$ be a polynomial in t with real coefficients. If the real part $\mathfrak{Re}(r)$ is negative for every root r of $p(t)$, then all the coefficients of $p(t)$ are positive.*

Proof. Let $a > 0$. If $-a < 0$ is a real root of $p(t)$, then $t + a$ is a factor of $p(t)$. If $-a + bi$ is a complex root of $p(t)$ for some $b \in \mathbb{R}$, then $-a - bi$ must be a root of $p(t)$ as well, which implies that

$$(t + a - bi)(t + a + bi) = (t^2 + 2at + a^2 + b^2)$$

is a factor of $p(t)$. Hence, $p(t)$ is a product of factors with positive coefficients. Thus, our conclusion follows. □

We say that a polynomial (with real coefficients) is *negative-real-part-rooted* or *NRPR* if all of its roots have negative real parts. The above lemma implies that if $i(P, t)$ is NRPR, then P is Ehrhart positive. Ehrhart polynomials of unit cubes and standard simplices are all trivially NRPR, as they factor into linear polynomials with positive real coefficients. Hence, we would like to rule them out and are only interested in examples of Ehrhart polynomials that are nontrivially NRPR.

It turns out that the following theorem which establishes a connection between roots of the h^*-polynomial and roots of the Ehrhart polynomial of a polytope is very useful.

Theorem 2.2.2 ([88], Theorem 3.2 of [101])**.** *Let P be a d-dimensional integral polytope, let k be the degree of the polynomial $h^*_P(z)$ (so $0 \le k \le d$), and suppose that every root of $h^*_P(z)$ lies on the circle $\{z \in \mathbb{C} \ : \ |z| = 1\}$ in the complex plane. Then there exists a polynomial $f(t)$ of degree k such that*

$$i(P, t) = f(t) \cdot \prod_{i=1}^{d-k} (t + i),$$

and every root of $f(t)$ has real part $-(1 + (d - k))/2$.

We say a polytope P is h^*-*unit-circle-rooted* if the h^*-polynomial $h^*_P(z)$ of P has all of its roots on the unit circle of the complex plane. Below we introduce three families of polytopes and show that each polytope P in these families is h^*-unit-circle-rooted. Therefore, Ehrhart positivity for these families follows from Theorem 2.2.2 and Lemma 2.2.1.

2.2.1 Cross-Polytopes

The d-dimensional cross-polytope, denoted by $\Diamond_d$, is a polytope in $\mathbb{R}^d$ defined by

$$\lozenge_d := \operatorname{conv}\{\pm \boldsymbol{e}_1, \pm \boldsymbol{e}_2, \ldots, \pm \boldsymbol{e}_d\},$$

or equivalently by the following linear system:

$$\pm x_1 \pm x_2 \pm \cdots \pm x_d \leq 1.$$

Hence, $i(\lozenge_d, t)$ counts the number of integer solutions to

$$|x_1| + |x_2| + \cdots |x_d| \leq t.$$

Counting the number of integer solutions with exactly k nonzero x_i's for $k = 0, 1, 2, \ldots, d,$ we obtain that

$$i(\lozenge_d, t) = \sum_{k=0}^{d} 2^k \binom{d}{k}\binom{t}{k}.$$

Unfortunately, it is *not* clear whether the above expression expands positively in powers of t. We compute its Ehrhart series instead. First, notice that $i(\lozenge_d, t)$ counts the number of integer solutions to

$$|x_1| + |x_2| + \cdots |x_d| + y = t.$$

Hence,

$$i(\lozenge_d, t) = \sum f(a_1) f(a_2) \ldots f(a_d)\, g(b),$$

where the summation is over all weak compositions $(a_1, \ldots, a_d, b)$ of t into $d+1$ parts, $g(b) = 1$ for all $b \geq 0$ and $f(a) = 1$ if $a = 0$ and $f(a) = 2$ if $a > 0$. Therefore,

$$\sum_{t \geq 0} i(\lozenge_d, t) z^t = \prod_{i=1}^{d} \left(\sum_{a_i \geq 0} f(a_i) z^{a_i} \right) \cdot \sum_{b \geq 0} z^b = \left(\frac{1+z}{1-z} \right)^d \cdot \frac{1}{1-z} = \frac{(1+z)^d}{(1-z)^{d+1}}.$$

Thus, $(1+z)^d$ is the h^*-polynomial of the cross-polytope $\lozenge_d$. Hence, $\lozenge_d$ is h^*-unit-circle-rooted and thus is Ehrhart positive.

2.2.2 Certain Families of $\boldsymbol{\Delta_{(1,q)}}$

Let $\boldsymbol{q} = (q_1, q_2, \ldots, q_d) \in \mathbb{P}^d$ be a sequence of positive integers. For each such a vector $\boldsymbol{q}$, we define a simplex

$$\Delta_{(1,\boldsymbol{q})} := \operatorname{conv}\left\{ \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_d, -\sum_{i=1}^{d} q_i \boldsymbol{e}_i \right\}.$$

In [29], Conrads studied simplices of this form and showed that $\Delta_{(1,\boldsymbol{q})}$ is reflexive if and only if

$$q_i \text{ divides } 1 + \sum_{j=1}^{d} q_j, \quad \text{for } i = 1, 2, \ldots, d.$$

Recently, Braun, Davis, and Solus studied $\Delta_{(1,\boldsymbol{q})}$ in their investigation of a conjecture by Hibi and Ohsugi, and they provided a number-theoretic characterization of the h^*-polynomial of $\Delta_{(1,\boldsymbol{q})}$ [15, Theorem 2.5].

Theorem 2.2.3 (Braun-Davis-Solus)**.** *The h^*-polynomial of $\Delta_{(1,\boldsymbol{q})}$ is*

$$h^*\left(\Delta_{(1,\boldsymbol{q})}, z\right) = \sum_{b=0}^{q_1+q_2+\cdots+q_d} z^{\omega(b)}, \tag{2.5}$$

where

$$\omega(b) := b - \sum_{i=1}^{d} \left\lfloor \frac{q_i b}{1 + q_1 + \cdots + q_d} \right\rfloor.$$

Formula (2.5) allows us to compute the h^*-polynomial for $\Delta_{(1,\boldsymbol{q})}$ with special choices of $\boldsymbol{q}$ easily. We give two examples below in which $\Delta_{(1,\boldsymbol{q})}$ satisfies the hypothesis of Theorem 2.2.2 with $k = d$.

Example 2.2.4 (Standard reflexive simplices)**.** If we choose $\boldsymbol{q} = \mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{P}^d$, then $\Delta_{(1,\boldsymbol{q})}$ is the d-dimensional *standard reflexive simplex*. Note that in this case, we have that $q_1 + q_2 + \cdots + q_d = d$. Furthermore, for each $b \in \{0, 1, 2, \ldots, d\}$, one can verify that $w(b) = b$. Hence,

$$h^*\left(\Delta_{(1,\mathbf{1})}, z\right) = \sum_{b=0}^{d} z^b = 1 + z + z^2 + \cdots + z^d.$$

Example 2.2.5 (Payne's construction)**.** In [78], Payne constructed reflexive simplices that do not have unimodal h^*-vectors. His construction is equivalent to the simplices $\Delta_{(1,\boldsymbol{q})}$ with the following choices of $\boldsymbol{q}$: Given $r \geq 0$, $s \geq 3$ and $k \geq r + 2$, let $d = r + sk$ and

$$\boldsymbol{q} = (q_1, q_2, \ldots, q_d) = (\underbrace{1, 1, \ldots, 1}_{sk-1 \text{ times}}, \underbrace{s, s, \ldots, s}_{r+1 \text{ times}}).$$

Applying Theorem 2.2.3, one can obtain

$$h^*\left(\Delta_{(1,\boldsymbol{q})}, z\right) = (1 + z^k + z^{2k} + \cdots + z^{(s-1)k})(1 + z + z^2 + \cdots + z^{k+r}).$$

Polytopes in both Examples 2.2.4 and 2.2.5 are h^*-unit-circle-rooted. Hence, they are Ehrhart positive. We remark that the Ehrhart positivity of $\Delta_{(1,\boldsymbol{q})}$ considered in Example 2.2.5 was first proved by the author and Solus [60, Theorem 3.2].

2.2.3 One Family of Order Polytopes

Given a finite poset (partially ordered set) $\mathcal{P}$, the *order polytope*, denoted by $\mathcal{O}(\mathcal{P})$, is the collection of functions $\boldsymbol{x} \in \mathbb{R}^{\mathcal{P}}$ satisfying

- $0 \le x_a \le 1$, for all $a \in \mathcal{P}$, and
- $x_a \le x_b$, if $a \le b$ in $\mathcal{P}$.

The order polytope $\mathcal{O}(\mathcal{P})$ was first defined and studied by Stanley [102]. Here we consider a family of order polytopes constructed from a certain family of posets.

Let $\mathcal{P}_k$ be the ordinal sum of k copies of 2 element antichains; equivalently, $\mathcal{P}_k$ is the poset on the $2k$-element set $\{a_{i,j} \ : \ i = 1, 2 \text{ and } j = 1, 2, \dots, k\}$ satisfying

$$a_{i,j} \le a_{i',j'} \text{ if and only if } j < j' \text{ or } (i, j) = (i', j').$$

For any $t \in \mathbb{N}$, the tth dilation $t\mathcal{O}(\mathcal{P}_k)$ of $\mathcal{O}(\mathcal{P}_k)$ is the collection of $\boldsymbol{x} = (x_{i,j} \ : \ i = 1, 2 \text{ and } j = 1, 2, \dots, k) \in \mathbb{R}^{2k}$ satisfying

$$0 \le x_{i,j} \le t, \quad \text{and} \quad x_{i,j} \le x_{i',j'} \text{ if } j < j'.$$

Hence, $i(\mathcal{O}(\mathcal{P}_k), t)$ counts the number of integer solutions $\boldsymbol{x}$ satisfying the above two conditions. Note that each solution gives a weak composition $(y_1, z_1, y_2, \dots, y_k, z_k, y_{k+1})$ of t into $2k + 1$ parts, where

$$\begin{aligned} y_j &= \min(x_{1,j}, x_{2,j}) - \max(x_{1,j-1}, x_{2,j-1}), \quad \text{for } j = 1, 2, \dots, k+1, \\ z_j &= \max(x_{1,j}, x_{2,j}) - \min(x_{1,j}, x_{2,j}) = |x_{1,j} - x_{2,j}|, \quad \text{for } j = 1, 2, \dots, k, \end{aligned}$$

and by convention let $\max(x_{1,0}, x_{2,0}) = 0$ and $\min(x_{1,k+1}, x_{2,k+1}) = t$. Thus,

$$i(\mathcal{O}(\mathcal{P}_k), t) = \sum g(y_1) f(z_1) g(y_2) f(z_2) \dots f(z_k) g(y_{k+1}),$$

where the summation is over all weak compositions of t into $2k + 1$ parts, $g(y) = 1$ for all $y \ge 0$, and $f(z) = 1$ if $z = 0$ and $f(z) = 2$ if $z > 0$. Therefore, similar to the calculation for cross-polytopes, we obtain

$$\sum_{t \ge 0} i(\mathcal{O}(\mathcal{P}_k), t) z^t = \frac{(1+z)^k}{(1-z)^{2k+1}}.$$

Thus, the h^*-polynomial of $\mathcal{O}(\mathcal{P}_k)$ is $(1 + z)^k$. By Lemma 2.2.1 and Theorem 2.2.2, the order polytope $\mathcal{O}(\mathcal{P}_k)$ is Ehrhart positive.

Remark 2.2.6. Stanley also defined a "chain order" polytope $\mathcal{C}(\mathcal{P})$ for each poset $\mathcal{P}$ [102, Definition 2.1] and showed that $\mathcal{C}(\mathcal{P})$ is unimodularly equivalent to $\mathcal{O}(\mathcal{P})$ [102, Theorem 3.2/(b)], from which it follows that $i(\mathcal{C}(\mathcal{P}), t) = i(\mathcal{O}(\mathcal{P}), t)$.

Therefore, the conclusions we draw above for the order polytope $\mathcal{O}(\mathcal{P}_k)$ all hold for the chain polytope $\mathcal{C}(\mathcal{P}_k)$.

It turns out that the polytopes studied in Sects. 2.2.1 and 2.2.2 are "reflexive" polytopes, and the order polytopes studied in Sect. 2.2.3 are "Gorenstein" polytopes. These are not coincidences as we will discuss below.

Connection to Reflexivity and Gorensteinness. An integral polytope P in $\mathbb{R}^D$ is *reflexive* (up to lattice translation) if the origin is in the interior of P and its *dual*

$$P^\vee := \{\boldsymbol{y} \in (\mathbb{R}^D)^* \ : \ \langle \boldsymbol{y}, \boldsymbol{x} \rangle \le 1 \ \forall \boldsymbol{x} \in P\}$$

is also an integral polytope, where $(\mathbb{R}^D)^*$ is the dual space of $\mathbb{R}^D$.

It follows from the Macdonald Reciprocity Theorem [64] that if an integral polytope P is reflexive, then the roots of $i(P, t)$ are symmetrically distributed with respect to the line $\{z \in \mathbb{C} \ : \ \Re e(z) = -1/2\}$ in the complex plane. Bey, Henk, and Wills show that the converse is true if we include polytopes that are unimodularly equivalent to reflexive polytopes [12, Proposition 1.8]. Recently, Hegedüs, Higashitani, and Kasprzyk, in their study of roots of Ehrhart polynomials of reflexive polytopes, give the following result [41, Lemma 1.2].

Lemma 2.2.7 (Hegedüs-Higashitani-Kasprzyk). *A d-dimensional integral polytope P is reflexive (up to unimodular transformation) if and only if the summation of the d roots of $i(P, t)$ equals to $-d/2$.*

Reflexive polytopes are special cases of a more general family of polytopes: Gorenstein polytopes. Recall that the *codegree* of P is defined to be

$$\operatorname{codeg}(P) := \dim(P) + 1 - \deg\left(h^*_P(z)\right).$$

It is (again) a consequence of the Macdonald Reciprocity Theorem [64] that $\operatorname{codeg}(P)$ is the smallest positive integer s such that sP contains a lattice point in its interior (see, e.g., [48]). A *Gorenstein polytope* is an integral polytope P of codegree s such that sP is a reflexive polytope. The work [95] of Stanley gives a nice characterization for the h^*-polynomials of Gorenstein polytopes: A d-dimensional integral polytope P is a Gorenstein polytope if and only if its h^*-polynomial is *symmetric*; that is, if $h^*_P(z) = \sum_{i=0}^k h^*_i z^i$ with $h^*_k \neq 0$, then $h^*_i = h^*_{k-i}$ for $i = 0, 1, 2, \ldots, k$. Using this, one can easily see that all the examples discussed in Sects. 2.2.1–2.2.3 are Gorenstein polytopes.

We now restate Lemma 2.2.7 in terms of Gorenstein polytopes.

Lemma 2.2.8. *A d-dimensional integral polytope P is Gorenstein (up to unimodular transformation) if and only if the summation of the d roots of $i(P,t)$ equals to $-sd/2$ for some positive integer s.*

Furthermore, if the above condition holds, the integer s is the codegree of P.

Proof. The conclusion of the lemma follows from the observation that a number t_0 is a root of $i(P,t)$ if and only if t_0/s is a root of $i(sP,t)$.

Corollary 2.2.9. *Suppose P is a d-dimensional polytope that is h^*-unit-circle-rooted. Then P is a Gorenstein polytope (up to unimodular transformation).*

Moreover, if the degree of the h^*-polynomial $h^*_P(z)$ is d, then P is reflexive.

Proof. Let k be the degree of $h^*_P(z)$. By Theorem 2.2.2, among all the roots of $i(P,t)$, k of them have real parts $-(1+(d-k))/2$, and the other $(d-k)$ roots are $-1, -2, \ldots, -(d-k)$. As $i(P,t)$ is a polynomial with real coefficients, the sum of roots of $i(P,t)$ is the sum of the real parts of roots of $i(P,t)$, which is

$$-(1+(d-k))/2 \cdot k + \sum_{i=1}^{d-k}(-i) = -\frac{1}{2}d(d-k+1).$$

Then the conclusions follow from Lemma 2.2.8. □

Therefore, when an integral polytope P is h^*-unit-circle-rooted, we not only get Ehrhart positivity for P but can also conclude that P is a Gorenstein polytope of codegree $d-k+1$, where k is the degree of the h^*-polynomial of P.

2.3 *Coefficients with Combinatorial Meanings*

2.3.1 Zonotopes

In this part, we introduce a special family of polytopes, *zonotopes*, whose Ehrhart coefficients can be described combinatorially. As a consequence, Ehrhart coefficients of a zonotope are not only positive but also positive integers.

The *Minkowski sum* of two polytopes (or sets) P and Q is

$$P + Q := \{\boldsymbol{x} + \boldsymbol{y} \ : \ \boldsymbol{x} \in P, \ \boldsymbol{y} \in Q.\}.$$

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \in \mathbb{Z}^D$ be a set of integer vectors. The *zonotope* $\mathcal{Z}(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ associated with this set of vectors is the Minkowski sum of intervals $[0, \boldsymbol{v}_i]$, where $[0, \boldsymbol{v}_i]$ is the line segment from the origin to $\boldsymbol{v}_i$. Hence,

$$\mathcal{Z}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) := \sum_{i=1}^n [0, \boldsymbol{v}_i] = \left\{ \sum_{i=1}^n c_i \boldsymbol{v}_i \ : \ 0 \le c_i \le 1 \text{ for } i = 1, 2, \ldots, n \right\}.$$

In [92, Theorem 2.2], Stanley gives a combinatorial description for the Ehrhart coefficients of zonotopes.

Theorem 2.3.1 (Stanley). *The coefficient of t^k in $i(\mathcal{Z}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n), t)$ is equal to*

$$\sum_X h(X),$$

where X ranges over all linearly independent k-subsets of $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$, and $h(X)$ is the greatest common divisor of all $k \times k$ minors of the matrix whose column vectors are elements of X.

The main ingredient for the proof of the above theorem is that $\mathcal{Z}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ can be written as a disjoint union of half-open parallelepiped C_X ranging over all linearly independent subsets $X = \{\boldsymbol{v}_{j_1}, \dots, \boldsymbol{v}_{j_k}\}$ of $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$, where C_X is generated by $\epsilon_1 \boldsymbol{v}_{j_1}, \dots, \epsilon_i \boldsymbol{v}_{j_k}$ for certain choices of $\epsilon_1, \dots, \epsilon_k \in \{-1, 1\}$. (see [92, Lemma 2.1].) The theorem then follows from the fact that the number of lattice points in the half-open parallelepiped C_X is the volume of C_X, which can be calculated by $h(X)$.

The simplest examples of zonotopes are unit cubes considered in Sect. 2.1.1. We may recover the Ehrhart polynomial of a unit cube using Theorem 2.3.1. However, a more interesting example is the *regular permutohedron*.

Example 2.3.2. The *regular permutohedron*, denoted by Π_d, is the convex hull of all permutations in $\mathfrak{S}_{d+1}$; that is,

$$\Pi_d := \operatorname{conv}\{(\sigma(1), \sigma(2), \dots, \sigma(d+1)) \in \mathbb{R}^{d+1} : \sigma \in \mathfrak{S}_{d+1}\}.$$

It is straightforward to check that Π_d is a translation of the zonotope

$$\sum_{1 \le i < j \le d+1} [0, \boldsymbol{e}_j - \boldsymbol{e}_i].$$

For any subset X of $\Phi_d := \{\boldsymbol{e}_j - \boldsymbol{e}_i \ : \ 1 \le i < j \le d+1\}$, we let G_X be the graph on vertex set $[d+1]$ and $\{i, j\}$ (with $i < j$) is an edge if and only if $\boldsymbol{e}_j - \boldsymbol{e}_i \in X$. Then it follows from matroid theory that X is linearly independent if and only if G_X is a forest on $[d+1]$. (Recall that a *forest* is a collection of trees or equivalently is an acyclic graph.) Furthermore, if X is linearly independent, then G_X is a forest of $d+1-|X|$ trees, and $h(X)$ (described in Theorem 2.3.1) is 1.

Therefore, we conclude that the coefficient of t^k in $i(\Pi_d, t)$ counts the number of forests on $[d+1]$ that contain exactly $d+1-k$ trees. Therefore, we can compute, for example,

$$i(\Pi_3, t) = 16t^3 + 15t^2 + 6t + 1.$$

2.3.2 Positivity of a Generalized Ehrhart Polynomial

The polynomial we discuss in this part is not exactly an Ehrhart polynomial. However, it is closely related to the result on zonotopes we have presented in the last part, and thus is included here. Galashin, Hopkins, McConville, and Postnikov, in their study of root system chip firing [38], considered the following lattice points counting problem: Given an integral polytope P and a set of integer vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$, describe the number of lattice point in

$$P + \boldsymbol{v}_1 + \boldsymbol{v}_2 + \cdots + \boldsymbol{v}_n = P + \mathcal{Z}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n).$$

Extending Stanley's idea of decomposing zonotopes into half-open parallelepiped, they show [38, Proof of Theorem 16.1] that $P + \mathcal{Z}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ can be written as disjoint union of sets in the form of $F + C_X$ where F is an open face of P and C_X is a half-open parallelepiped determined by a linearly independent set X of $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$.

Theorem 2.3.3 (Galashin-Hopkins-McConville-Postnikov). *Suppose P is an integral polytope in $\mathbb{R}^D$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{Z}^D$ is a set of integer vectors. Let $\mathcal{Z} = \mathcal{Z}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$. For any $\boldsymbol{t} = (t_1, \ldots, t_n) \in \mathbb{N}_n$, we define $\boldsymbol{t}\mathcal{Z} = \mathcal{Z}(t_1\boldsymbol{v}_1, \ldots, t_n\boldsymbol{v}_n)$. Then there exists a polynomial $L(\boldsymbol{t}) = L(t_1, \ldots, t_n)$ in n variables with nonnegative integer coefficients such that $|(P + \boldsymbol{t}\mathcal{Z}) \cap \mathbb{Z}^D| = L(\boldsymbol{t})$.*

In particular, if we take $\boldsymbol{t} = (t, t, \ldots, t)$, then

$$|(P + \boldsymbol{t}\mathcal{Z}) \cap \mathbb{Z}^D| = |(P + t\mathcal{Z}) \cap \mathbb{Z}^D| = L(t, t, \ldots, t)$$

is a polynomial in t of degree $\dim(\mathcal{Z})$ *with positive integer coefficients.*

Note that the second part of the above theorem was not explicitly stated in [38, Theorem 16.1], but it was a consequence of the techniques used in its proof.

One sees that if we choose P to be the origin, then the above theorem recovers the Ehrhart positivity of zonotopes. However, in contrast with Stanley's results, no explicit formulas were given in [38] for the positive/nonnegative integer coefficients asserted in Theorem 2.3.3. Recently, Hopkins and Postnikov [49] analyzed techniques used in [38] further and provided the desired explicit formula, completing the generalization of Theorem 2.3.1.

Theorem 2.3.4 (Hopkins-Postnikov). *The homogeneous degree k part of the polynomial $L(\boldsymbol{t})$ assumed by Theorem 2.3.3 is given by*

$$\sum_X |\operatorname{quot}_X(P) \cap \operatorname{quot}_X(\mathbb{Z}^D)| \cdot h(X) \cdot \prod_{\boldsymbol{v}_i \in X} t_i,$$

where X ranges over all linearly independent k-subsets of $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$, $\operatorname{quot}_X : \mathbb{R}^D \to \mathbb{R}^D/\operatorname{span}_{\mathbb{R}}(\mathrm{X})$ *is the canonical quotient map, and $h(X)$ is the greatest common divisor of all $k \times k$ minors of the matrix whose column vectors are elements of X.*

2.4 Higher Integrality Conditions

In this part, we will introduce families of polytopes whose Ehrhart coefficients are always volumes of certain projections of the original polytopes and are hence positive.

2.4.1 Cyclic Polytopes

We start with a well-known family of polytopes: *cyclic polytopes*. The *moment curve* in $\mathbb{R}^d$ is defined by

$$\nu_d : \mathbb{R} \to \mathbb{R}^d, x \mapsto \nu_d(u) = (u, u^2, \dots, u^d).$$

Let $U = \{u_1, \dots, u_n\}_<$ be a linear ordered set. Then the *cyclic polytope* $C_d(U) = C_d(u_1, \dots, u_n)$ is the convex hull of $n > d$ distinct points $\nu_d(t_i)$, $1 \le i \le n$, on the moment curve:

$$C_d(U) := \text{conv}\{\nu_d(u_1), \nu_d(t_2), \dots, \nu_d(u_n)\}.$$

Cyclic polytopes form an interesting family of polytopes. For instance, its facets are determined by the Gale evenness condition [108, Theorem 0.7], and the number of i-dimensional faces of $C_d(U)$ (where $|U| = n$) is the upper bound for the number of i-dimensional faces of all d-dimensional polytopes with n vertices [66].

The following theorem on the Ehrhart polynomial of integral cyclic polytopes was initially conjectured in [9] by Beck, De Loera, Develin, Pfeifle, and Stanley and then proved in [57] by the author.

Theorem 2.4.1 (L.). *For any d-dimensional integral cyclic polytope $P = C_d(U) \subset \mathbb{R}^d$, we have that*

$$i(P,t) = \text{Vol}_d(P)t^d + i(\pi(P),t) = \sum_{k=0}^{d} \text{Vol}_k\left(\pi^{(d-k)}(P)\right)t^k, \tag{2.6}$$

where $\pi^{(d-k)} : \mathbb{R}^d \to \mathbb{R}^k$ is the map that ignores the last $d-k$ coordinates of a point, and $\text{Vol}_k(Q)$ *is the volume of Q in the k-dimensional space $\mathbb{R}^k$.*

The first step of the proof is to reduce the problem to simplices by using triangulations. For the simplex case, we consider the set obtained by removing the lower envelope of $C_d(U)$ (with $|U| = d + 1$), and we decompose this set into $d!$ signed (convex) half-open sets S_σ, each of which corresponds to a permutation σ in the symmetric group $\mathfrak{S}_d$. One important feature of this decomposition is that the number of lattice points in each piece S_σ can be expressed in a simple formula involving the permutation σ, which makes it possible to compute the summation of all $d!$ terms.

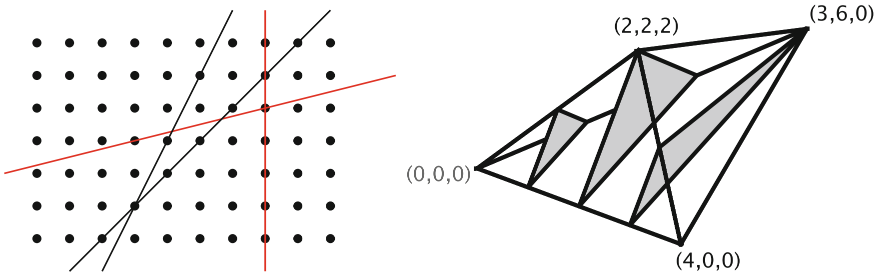


Fig. 1 Examples of higher integrality conditions

2.4.2 *k*-Integral Polytopes

Since the work in [57], the author generalized the family of integral cyclic polytopes to a bigger family of integral polytopes, "lattice-face polytopes," and showed that their Ehrhart polynomials are also in the simple form of (2.6) [56, 58]. Later in [59], the author improved her results by introducing a notion of "higher integrality," which we will detail below.

Recall that a lattice point is also called an *integral* point. A point can be considered as a 0-dimensional affine space. We first extend this concept of integrality to higher-dimensional affine spaces: An ℓ-dimensional affine space W in $\mathbb{R}^d$ is *integral* if

$$\pi^{(d-\ell)}(W \cap \mathbb{Z}^d) = \mathbb{Z}^\ell.$$

Note that this definition with $\ell = 0$ is consistent with the original definition of an integral point.

Example 2.4.2 (lines in $\mathbb{R}^2$)**.** See the left side of Fig. 1 for examples of 1-dimensional affine space in $\mathbb{R}^2$. The black lines are integral, while the red lines are *not* integral. For the slanted red line, say L_1, we have $\pi^{(2-1)}(L_1 \cap \mathbb{Z}^2) \cong \mathbb{Z}/4\mathbb{Z}$. For the vertical red line, say L_2, we have $\pi^{(2-1)}(L_2 \cap \mathbb{Z}^2) \cong \mathbb{Z}^0$.

Note that in the above example, even though L_1 is not integral, after the projection, we still get a 1-dimensional lattice, which has the same dimension as L_1. In this case, we say L_1 is *in general position*. On the contrary, L_2 is *not* in general position.

Definition 2.4.3. Suppose $0 \le k \le d$. A d-dimensional polytope P is *k-integral* if for any face F of P of dimension less than or equal to k, the affine hull $\operatorname{aff}(F)$ of F is integral.

In particular, when $k = d$, we call P a *fully integral* polytope.

Remark 2.4.4. With the above definition, *lattice-face polytopes*, introduced in [56, 58], can be defined as polytopes that can be triangulated into fully integral simplices, which is a property any (integral) cyclic polytope has. Therefore, any cyclic polytope or lattice polytope is fully integral.

The main result in [59] is a complete description for the Ehrhart coefficients of a k-integral polytope in terms of volumes of projections and Ehrhart polynomials of *slices*.

Definition 2.4.5. For any $\boldsymbol{y} \in \pi^{(d-k)}(P)$, we define *the slice of P over $\boldsymbol{y}$*, denoted by $\pi_{d-k}(\boldsymbol{y}, P)$, to be the intersection of P with the inverse image of $\boldsymbol{y}$ under $\pi^{(d-k)}$.

Recall that $[t^k]f(t)$ denotes the coefficient of t^k of a polynomial $f(t)$.

Theorem 2.4.6 (L.). *If P is a k-integral polytope, then*

$$[t^\ell]i(P,t) = \begin{cases} \mathrm{Vol}(\pi^{d-\ell}(P)) & \text{if } 0 \le \ell \le k, \\ [t^{\ell-k}]\left(\sum_{\boldsymbol{y}\in\pi^{(d-k)}(P)\cap\mathbb{Z}^k} i(\pi_{d-k}(\boldsymbol{y},P),t)\right) & \text{if } k+1 \le \ell \le d. \end{cases}$$

Therefore, if P is fully integral, the Ehrhart polynomial of P is in the form of (2.6)*, and thus P is Ehrhart positive.*

Because both cyclic polytopes and lattice-face polytopes are fully integral polytopes, the above theorem generalizes results in [56–58].

The following is an example showing how to use Theorem 2.4.6 to obtain the Ehrhart polynomial of a 1-integral polytope.

Example 2.4.7 (Example of Theorem 2.4.6)**.** Consider the 3-dimensional polytope

$$P = \mathrm{conv}\{(0,0,0),(4,0,0),(3,6,0),(2,2,2)\} \subset \mathbb{R}^3,$$

which is illustrated on the right side of Fig. 1. One checks that P is 1-integral. Clearly $\pi^{(2)}(P) = [0,4]$ and $\pi^{(3)}(P) = 0$. By the first part of Theorem 2.4.6,

$$[t^1]i(P,t) = \mathrm{Vol}_1([0,4]) = 4, \quad \text{and} \quad [t^0]i(P,t) = \mathrm{Vol}_0(0) = 1.$$

For the higher Ehrhart coefficients of P, we need to compute the Ehrhart polynomials of slices of P over lattice points in $\pi^{(2)}(P) = [0,4]$. In the picture, the three shaded triangles are the slices of P over the lattice points 1, 2, and 3. The slices of P over lattice points 0 and 4 are the single points $(0,0,0)$ and $(4,0,0)$, respectively. We calculate the Ehrhart polynomials of all five slices, by summing which up we obtain $8t^2 + 10t + 5$. Then the second part of Theorem 2.4.6 says that

$$[t^3]i(P,t) = 8 \quad \text{and} \quad [t^2]i(P,t) = 10.$$

Therefore,

$$i(P,t) = 8t^3 + 10t^2 + 4t + 1.$$

Recall that the *face poset* of a polytope P is the set of all faces of P ordered by inclusion. We say two polytopes have the *same combinatorial type* if they have

the same face poset. As a by-product of the study of Ehrhart polynomials of full-integral polytopes, we can also show that Ehrhart positivity is independent from combinatorial types of polytopes [56].

Theorem 2.4.8 (L.). *For any rational polytope P, there exists a polytope P' with the same face lattice such that P' satisfies the higher integrality condition and thus is Ehrhart positive.*

Sketch of proof. First, by choosing appropriate bases for our underlying lattice $\mathbb{Z}^d$, we may assume that the affine hull of any face of P is in general position.

Next, for any $\boldsymbol{s} = (s_1, \dots, s_d) \in \mathbb{Z}^d$ and $\boldsymbol{x} \in \mathbb{R}^d$, we define

$$\boldsymbol{s} \star \boldsymbol{x} = (s_1 x_1, s_2 x_2, \dots, s_d x_d).$$

So $\boldsymbol{s}$ is an operator on $\mathbb{R}^d$ that dilates points with different scalars at different coordinates. We observe that for any ℓ-dimensional affine space $W \subset \mathbb{R}^d$ that is in general position, there exist (positive) integer scalars $c_1, \dots, c_\ell$ such that for any $\boldsymbol{s} \in \mathbb{Z}^d_{\neq 0}$, if $c_m s_m$ divides s_{m+1} for each $m \in \{1, 2, \dots, \ell\}$, then

$$\boldsymbol{s} \star W := \{\boldsymbol{s} \star \boldsymbol{w} \ : \ \boldsymbol{w} \in W\}$$

is integral. For example, for the slanted red line L_1 appeared in Example 2.4.2, one checks that whenever $\boldsymbol{s} = (s_1, s_2)$ satisfies $4s_1$ divides s_2, the affine space $\boldsymbol{s} \star L_1$ is integral. Hence, we can choose $c_1 = 4$.

Since P has finitely many faces, we can apply the above operations inductively on dimensions of faces to obtain a full-integral polytope P' that actually defined as $\boldsymbol{s} \star P$ for some $\boldsymbol{s} \in \mathbb{Z}^d_{\neq 0}$. □

Remark 2.4.9. There are a lot of properties of polytopes people study other than Ehrhart positivity, such as "normality," "integer decomposition property" (or IDP), and "existence of a (regular) unimodular triangulation." For the majority of them, even if you start with a polytope P that does not have a certain property, dilating P with a large enough scalar often yields a polytope with the desired property (see, e.g., [19, 30, 39]). Clearly, simple dilations would not change the answer to the Ehrhart positivity question for any polytope. After all, $i(kP, t) = i(P, kt)$. Hence, the Ehrhart coefficients of a dilation of P have exactly the same sign pattern as Ehrhart coefficients of P.

However, our proof of Theorem 2.4.8 says that dilating in different directions with different scalars can change a non-Ehrhart-positive polytope to a Ehrhart-positive one.

3 McMullen's Formula and Positivity of Generalized Permutohedra

The main purpose of this section is to study the Ehrhart positivity conjecture for generalized permutohedra. After reviewing previously known results supporting this conjecture, we introduce *McMullen's formula*, which is a formula for computing the number of lattice points inside polytopes. This provides us a way of attacking the question of Ehrhart positivity by reducing the problem to "α-positivity." We then discuss the author's joint work [24, 27] with Castillo on the Ehrhart positivity conjecture of generalized permutohedra using this approach.

3.1 Motivation and Evidence

In this part, we discuss the motivation for considering the Ehrhart positivity conjecture of generalized permutohedra and prior work by Postnikov which provides evidence for this conjecture. We start by formally defining *generalized permutohedra*, the main family of polytopes we study in this section. Whenever we talk about generalized permutohedra, we have $D = d + 1$.

3.1.1 Definition and First Positivity Conjecture

Given a strictly increasing sequence $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{d+1}) \in \mathbb{R}^{d+1}$, we define the *usual permutohedron* associated with $\boldsymbol{\alpha}$ as

$$\mathrm{Perm}(\boldsymbol{\alpha}) := \mathrm{conv}\left((\alpha_{\pi(1)}, \alpha_{\pi(2)}, \cdots, \alpha_{\pi(d+1)}) \ : \ \pi \in \mathfrak{S}_{d+1}\right)$$

In particular, if $\boldsymbol{\alpha} = (1, 2, \ldots, d + 1)$, we obtain the regular permutohedron Π_d considered in Example 2.3.2. In [82], Postnikov defined *generalized permutohedra* to be polytopes that can be obtained from usual permutohedra by moving vertices while preserving all edge directions. (Note that in this definition, edges are allowed to degenerate, and hence vertices can collapse.)

In [31], De Loera, Haws, and Koeppe study the Ehrhart polynomials of matroid base polytopes, and conjecture those all have positive coefficients. However, it turns out that every matroid base polytope is a generalized permutohedron [1, Section 2]. In [24, 27], Castillo and the author generalize the conjecture of De Loera et al. to all integral generalized permutohedra:

Conjecture 3.1.1 (Castillo-L.). *All integral generalized permutohedra are Ehrhart positive.*

Indeed, due to Postnikov's work, a big family of generalized permutohedra is already known to be Ehrhart positive, which provides a strong evidence to the above conjecture. We describe his work below.

3.1.2 Ehrhart Positivity of Type-$\mathcal{Y}$ Generalized Permutohedra

In [82], Postnikov considers Minkowski sums of dilated simplices: For any nonempty subset $I \subseteq [d+1]$, define the simplex

$$\Delta_I := \operatorname{conv}\{\boldsymbol{e}_i \ : \ i \in I\}.$$

Let $\boldsymbol{y} = (y_I \ : \ \emptyset \neq I \subseteq [d+1]) \in \left(\mathbb{R}_{\geq 0}\right)^{2^{d+1}-1}$ be a vector indexed by nonempty subsets of $[d+1]$ with nonnegative entries. We define the polytope

$$P_d^{\mathcal{Y}}(\boldsymbol{y}) := \sum_{\emptyset \neq I \subseteq [d+1]} y_I \Delta_I$$

as the Minkowski sum of the simplices Δ_I dilated by the factor y_I. Postnikov shows that $P_d^{\mathcal{Y}}(\boldsymbol{y})$ is always a generalized permutohedron [82, Proposition 6.3]; however, not every generalized permutohedron can be expressed as $P_d^{\mathcal{Y}}(\boldsymbol{y})$ for some $\boldsymbol{y}$ [82, Remark 6.4]. Therefore, we call $P_d^{\mathcal{Y}}(\boldsymbol{y})$ a *type-$\mathcal{Y}$ generalized permutohedron*.

Postnikov then reformulates the construction of $P_d^{\mathcal{Y}}(\boldsymbol{y})$ using bipartite graphs: Let G be a subgraph of the bipartite graph $K_{c,d+1}$ without isolated vertices. Label the vertices of G on the left by $l_1, l_2, \dots, l_c$ and vertices on the right by $r_1, r_2, \dots, r_{d+1}$. For each $1 \leq j \leq c$, we let

$$I_j^G = \{i \in [d+1] \ : \ \{l_j, r_i\} \text{ is an edge of } G\}.$$

For any $(y_1, y_2, \dots, y_c) \in \left(\mathbb{R}_{\geq 0}\right)^c$, we define the polytope

$$P_G(y_1, \dots, y_c) := \sum_{j=1}^{c} y_j \Delta_{I_j}^G.$$

Remark 3.1.2. It is clear that $P_G(y_1, y_2, \dots, y_c)$ is the type-$\mathcal{Y}$ generalized permutohedron $P_d^{\mathcal{Y}}(\boldsymbol{y})$ where $y_I = \sum_{j:I_j=I} y_j$. Conversely, the type-$\mathcal{Y}$ generalized permutohedron $P_d^{\mathcal{Y}}(\boldsymbol{y})$ is the polytope $P_G(\boldsymbol{y})$ where G is the subgraph of $K_{2^{d+1}-1,d+1}$ such that left vertices of G are indexed by nonempty subsets I of $[d+1]$, and the left vertex l_I is adjacent to the right vertex r_i if and only if $i \in I$.

In [82, Section 11], Postnikov defines the "trimmed generalized permutohedron" as the "Minkowski difference" of $P_G(y_1, \dots, y_c)$ and the simplex $\Delta_{[d+1]}$. By providing a formula for the number of lattice points in a trimmed generalized permutohedron, he obtains a formula for the number of lattice points in $P_G(y_1, \dots, y_c)$

[82, Theorem 11.3], which leads to an expression for the Ehrhart polynomial of $P_G(y_1, \dots, y_c)$ as a summation over *G-draconian sequences*.

Definition 3.1.3 (Definition 9.2 in [82]). A sequence of nonnegative integers $\boldsymbol{g} = (g_1, g_2, \dots, g_c)$ is a *G-draconian sequence* if $\sum_{j=1}^{c} g_j = d$ and for any subset $\{j_1, \dots, j_k\} \subseteq [c]$, we have $|I_{j_1}^G \cup \cdots \cup I_{j_k}^G| \geq g_{j_1} + \cdots + g_{j_k} + 1$.

Theorem 3.1.4 (Postnikov). *Suppose G is a subgraph of $K_{c,d+1}$ without isolated vertices such that $I_1^G = [d+1]$. Let $y_1, \dots, y_c \in \mathbb{N}$. Then the Ehrhart polynomial of $P_G(y_1, \dots, y_c)$ is given by*

$$i(P_G(y_1, y_2, \dots, y_c), t) = \sum_{\boldsymbol{g}} \left(\!\!\binom{y_1 t + 1}{g_1}\!\!\right) \prod_{k=2}^{c} \left(\!\!\binom{y_k t}{g_k}\!\!\right),$$

where the summation is over all G-draconian sequences $\boldsymbol{g} = (g_1, \dots, g_c)$.

Similar to the results discussed in Sects. 2.1.3 and 2.1.4, it follows from Lemma 2.1.1/(b) that the Ehrhart polynomial described in the above theorem has positive coefficients. Thus, by Remark 3.1.2, we immediately have the following:

Corollary 3.1.5. *Any integral type-$\mathcal{Y}$ generalized permutohedron is Ehrhart positive.*

Note that as we pointed out above, type-$\mathcal{Y}$ generalized permutohedra do not contain all generalized permutohedra. Thus, Conjecture 3.1.1 does not follow from the above result.

Example 3.1.6 (Pitman-Stanley polytopes again). Let G be a subgraph of $K_{d+1,d+1}$ where for each $j \in [d+1]$, the left vertex l_j is adjacent to right vertices $r_j, r_{j+1}, \dots, r_{d+1}$. Then for any $\boldsymbol{y} = (y_1, \dots, y_{d+1}) \in \mathbb{N}^{d+1}$,

$$P_G(\boldsymbol{y}) = P_G(y_1, \dots, y_{d+1}) = \sum_{j=1}^{d+1} y_j \Delta_{[j,d+1]},$$

where $[j, d+1] = \{j, j+1, \dots, d+1\}$. It follows from Proposition 6.3 of [82] that the inequality description of this polytope is

$$P_G(\boldsymbol{y}) = \left\{ \boldsymbol{x} \in \mathbb{R}^{d+1} \ : \ x_i \geq 0 \text{ and } \sum_{j=1}^{i} x_j \leq \sum_{j=1}^{i} y_j \text{ for } i = 1, 2, \dots, d-1, \text{ and } \sum_{j=1}^{d+1} x_j = \sum_{j=1}^{d+1} y_j \right\}.$$

It is easy to see the map $\pi : \mathbb{R}^{d+1} \to \mathbb{R}^d$ that ignores the last coordinate of a point induces a unimodular transformation from $P_G(\boldsymbol{y})$ to the Pitman-Stanley polytope $\mathcal{PS}_d(\widehat{\boldsymbol{y}})$ considered in Sect. 2.1.3, where $\widehat{\boldsymbol{y}} = (y_1, y_2, \dots, y_d)$.

One can also check that the G-draconian sequences for the graph G given in this example are those $\boldsymbol{g} = (g_1, \dots, g_{d+1}) \in \mathbb{N}^{d+1}$ satisfying

$$\sum_{j=1}^{d} g_j = d, \; g_{d+1} = 0, \text{ and } \sum_{j=1}^{k} g_j \geq k \text{ for } k = 1, 2, \ldots, d-1.$$

Hence, it can be verified that Theorem 2.1.2 is a special case of Theorem 3.1.4.

The family of type-$\mathcal{Y}$ generalized permutohedra not only includes the Pitman-Stanley polytope as we have seen in the example above, but also includes associahedra, cyclohedra, and more (see [82, Section 8]). However, it follows from work by Ardila, Benedetti, and Doker that type-$\mathcal{Y}$ generalized permutohedra do not contain all matroid base polytopes [1, Proposition 2.3 and Example 2.6]. Therefore, Corollary 3.1.5 does not settle either Conjecture 3.1.1 or the Ehrhart positivity conjecture on matroid base polytopes by De Loera et al. [31].

3.2 *McMullen's Formula, α-Positivity, and a Reduction Theorem*

The goal of this part is to introduce McMullen's formula and discuss why it is a good tool to show Ehrhart positivity of a family of polytopes constructed from a fixed projective fan when an α-construction satisfies certain valuation properties. (We will discuss in Sect. 3.3 that generalized permutohedra form a family of polytopes constructed from the Braid fan. Hence, the techniques introduced here are relevant to our question.)

Throughout the rest of this section, we let V be a subspace of $\mathbb{R}^D$ and V^* be the dual space of V. For any polytope P, we use the notation $\mathrm{lin}(P)$ to denote the linear space obtained by shifting the affine span $\mathrm{aff}(P)$ of P to the origin.

3.2.1 Cones

We need the concepts of *cones*, particularly *feasible cones* and *normal cones*, before we start our discussion.

A *(polyhedral) cone* is the set of all nonnegative linear combinations of a finite set of vectors. A cone is *pointed* if it does not contain a line.

Definition 3.2.1. Suppose P is a polytope satisfying $\mathrm{lin}(P) \subseteq V$.

(i) The *feasible cone* of P at F is:

$$\mathrm{fcone}(F, P) := \{\boldsymbol{u} \in V : x + \delta\boldsymbol{u} \in P \text{ for sufficiently small } \delta\},$$

where x is any relative interior point of F. (It can be checked that the definition is independent of the choice of x.)

The *pointed feasible cone* of P at F is

$$\text{fcone}^p(F, P) = \text{fcone}(F, P)/\text{lin}(F).$$

(ii) Given any face F of P, the *normal cone* of P at F with respect to V is

$$\text{ncone}_V(F, P) := \left\{ \boldsymbol{u} \in V^* \ : \ \langle \boldsymbol{u}, \boldsymbol{p}_1 \rangle \geq \langle \boldsymbol{u}, \boldsymbol{p}_2 \rangle, \quad \forall \boldsymbol{p}_1 \in F, \quad \forall \boldsymbol{p}_2 \in P \right\}.$$

Therefore, $\text{ncone}_V(F, P)$ is the collection of linear functionals $\boldsymbol{u}$ in V^* such that $\boldsymbol{u}$ attains maximum value at F over all points in P.

The *normal fan* $\Sigma_V(P)$ of P with respect to V is the collection of all normal cones of P.

Normal cones and pointed feasible cones are related by polarity.

Definition 3.2.2. Let $K \subseteq V^*$ be a cone, and let W be the subspace of V^* spanned by K. (So W^* is a quotient space of V.) The *polar cone* of K is the cone

$$K^\circ = \{ \boldsymbol{y} \in W^* \ : \ \langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq 0, \ \forall \boldsymbol{x} \in K \}.$$

Lemma 3.2.3 (Lemma 2.4 of [24]). *Suppose P is a polytope satisfying* $\text{lin}(P) \subseteq V$ *and F is a face of P. Then* $(\text{ncone}_V(F, P))^\circ$ *is a pointed cone, and is invariant under the choice of V. So we may omit the subscript V and just write* $(\text{ncone}(F, P))^\circ$. *Furthermore,*

$$\text{ncone}(F, P)^\circ = \text{fcone}^p(F, P).$$

3.2.2 McMullen's Formula and a Refinement of Positivity

In 1975 Danilov asked, in the context of toric varieties, whether it is possible to construct a function α such that for any integral polytope P, we have

$$|P \cap \mathbb{Z}^D| = \sum_{F: \text{ a face of } P} \alpha(F, P) \, \text{nvol}(F), \tag{3.1}$$

where $\alpha(F, P)$ depends only on the normal cone of P at F, and $\text{nvol}(F)$ is the volume of F normalized to the lattice $\text{aff}(F) \cap \mathbb{Z}^D$.

McMullen [67] was the first to confirm the existence of (3.1) in a nonconstructive way. Hence, we refer to the above formula as *McMullen's formula*. Pommersheim and Thomas [81] provide a canonical construction based on choices of flags. Berline and Vergne [11] give a construction in a computable way. Most recently, Ring and Schürmann [87] give another construction for McMullen's formula based on a choice of fundamental cells.

Before discussing a specific construction, even the existence of McMullen's formula has interesting consequences. In fact, it was one of the ingredients used in proving the results on higher integrality conditions discussed in Sect. 2.4.2. More importantly, it provides another proof for Ehrhart's theorem (Theorem 1.1) as well as a refinement of Ehrhart positivity. Note that pointed feasible cones do not change

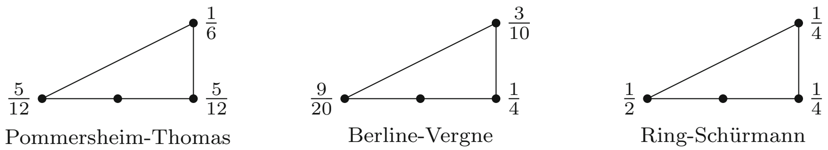


Fig. 2 Different α-constructions

when we dilate a polytope. Thus, applying McMullen's formula to tP and rearranging coefficients, we obtain a formula for the function $i(P, t)$:

$$i(P,t) = \sum_{k=0}^{\dim P} \left(\sum_{F:k\text{-dimensional face of } P} \alpha(F, P)\mathrm{nvol}(F) \right) \cdot t^k.$$

Hence, $i(P, t)$ is a polynomial in t of degree $\dim P$, and the coefficient of t^k in $i(P, t)$ is given by

$$[t^k]i(P,t) = \sum_{F:k\text{-dimensional face of } P} \alpha(F, P)\mathrm{nvol}(F). \tag{3.2}$$

Example 3.2.4. Setting $k = 0$ in (3.2), we obtain

$$[t^0]i(P,t) = \sum_{v:\text{vertex of } P} \alpha(v, P)\mathrm{nvol}(v).$$

Note that $[t^0]i(P, t)$ is the constant term of the Ehrhart polynomial of P, which is known to be 1 for any integral polytope P. Furthermore, the normalized volume of any vertex is 1. Hence, the above equation becomes

$$\sum_{v:\text{vertex of } P} \alpha(v, P) = 1.$$

See Fig. 2 for α-values of the vertices of the triangle $P = \mathrm{conv}((0, 0), (2, 0), (2, 1))$ arising from different constructions.

Since $\mathrm{nvol}(F)$ is a positive number, it follows from (3.2) that α-values refine Ehrhart coefficients. We say a polytope P is *α-positive for k-faces* if $\alpha(F, P)$ is positive for all k-dimensional faces F of P and say P is *α-positive* if all α's associated to P are positive. The following result immediately follows from expression (3.2).

Lemma 3.2.5. *Suppose α is a solution to McMullen's formula. Let P be an integral polytope. For a fixed k, if P is α-positive for k-faces, then the coefficient of t^k in the Ehrhart polynomial $i(P, t)$ of P is positive.*

Hence, if P is α-positive, then P is Ehrhart positive.

3.2.3 BV-Construction and the Reduction Theorem

At the first glance, α-positivity, being a refinement of Ehrhart positivity, is a more difficult question to consider. However, for α-constructions that satisfy certain properties, studying α-positivity instead does not necessarily make the problem harder. Berline and Vergne [11] give such an α-construction, of which we give a quick review below. Recall that the *indicator function* of a set $A \subseteq V$ is the function $[A] : V \to \mathbb{R}$ defined as $[A](x) = 1$ if $x \in A$, and $[A](x) = 0$ if $x \notin A$. The *algebra of rational cones*, denoted by $\mathcal{C}(V)$, is the vector space over $\mathbb{Q}$ spanned by the indicator functions $[C]$ of all rational cones $C \subset V$. We consider $\mathcal{C}(V)$ a subspace of the vector space of all functions on V. Hence, in general, the indicators $[C]$ of rational cones do not form a basis of $\mathcal{C}(V)$ since there are many relations among them.

Theorem 3.2.6 (Berline-Vergne). *There exists a function Ψ from the set of indicator functions $[C]$ of rational cones C in V to $\mathbb{R}$ with the following properties:*

(P1) *Ψ induces a* valuation *on the algebra of rational cones in V, i.e., Ψ induces a linear transformation from $\mathcal{C}(V)$ to $\mathbb{R}$.*
(P2) *If a cone C contains a line, then $\Psi([C]) = 0$.*
(P3) *Ψ is invariant under orthogonal unimodular transformation, thus, is* symmetric about coordinates*, that is, invariant under rearranging coordinates with signs.*
(P4) *Setting*

$$\alpha(F, P) := \Psi([\text{fcone}^p(F, P)]), \tag{3.3}$$

gives a solution to McMullen's formula.

We refer to Berline-Vergne construction of Ψ and α as *BV-construction* and *BV-α-valuation*, respectively. If α is the BV-α-valuation, we use the terminology *BV-α -positivity* instead of α-positivity.

Properties (P1) and (P2) are the "certain valuation properties" we mentioned at the beginning of Sect. 3.2. The following Reduction Theorem lays out a consequence of these two properties.

Theorem 3.2.7 (Castillo-L., Reduction Theorem [24]). *Suppose Ψ is a function from the set of indicator functions of rational cones C in V to $\mathbb{R}$ such that properties (P1) and (P2) hold, and suppose α is defined as in* (3.3).

Let P and Q be two polytopes such that $\text{lin}(P)$ *and* $\text{lin}(Q)$ *are both subspaces of V. Assume the normal fan $\Sigma_V(P)$ of P with respect to V is a refinement of the normal fan $\Sigma_V(Q)$ of Q with respect to V.*

Then for any fixed k, if P is α-positive for k-faces, then Q is α-positive for k-faces.

One important implication of the Reduction Theorem is that we can reduce the problem of α-positivity of a family of polytopes constructed from a fan to the problem of α-positivity of a single polytope in the family.

Definition 3.2.8. Let Σ be a *projective* fan in V^*, i.e., a fan that is a normal fan of some polytope. Let $\mathrm{Poly}(\Sigma)$ be the set of polytopes Q whose normal fan $\Sigma_V(Q)$ with respect to V coarsens Σ.

Corollary 3.2.9. *Assume the hypothesis on Ψ and α in Theorem 3.2.7. Let Σ be a projective fan in V^*, and let P be a polytope such that $\Sigma_V(P) = \Sigma$. Then α-positivity (for k-faces) of P implies α-positivity (for k-faces) of Q for any $Q \in \mathrm{Poly}(\Sigma)$.*

Assume further that α is a solution to McMullen's formula. Then for any integral polytope $Q \in \mathrm{Poly}(\Sigma)$, α-positivity for k-faces of P implies the coefficient of t^k in $i(Q, t)$ is positive. Hence, α-positivity of P implies Ehrhart positivity of Q.

Proof. The first part follows directly from the Reduction Theorem, and the second assertion follows from the first part and Lemma 3.2.5.

Therefore, even though proving α-positivity is more difficult than proving Ehrhart positivity for an individual polytope, it could be easier if we consider a family of polytopes $\mathrm{Poly}(\Sigma)$ associated to a fixed projective fan Σ, as we only need to prove α-positivity for one polytope in the family. Finally, because the BV-construction satisfies properties (P1), (P2), and (P4), all the results discussed above apply to the BV-construction or the BV-α-valuation. These ideas are illustrated by Example 3.3.3 below.

3.3 *Positivity of Generalized Permutohedra*

In this part, we apply the Reduction Theorem to reduce our first conjecture—Conjecture 3.1.1—to a conjecture on α-positivity of regular permutohedra. Then we report partial progress made on both conjectures by using McMullen's formula with BV-α-valuation [24, 27].

3.3.1 Second Positivity Conjecture

Postnikov, Reiner, and Williams give several equivalent definitions for generalized permutohedra, one of which uses concepts of normal fans [83, Proposition 3.2]. Recall that the *Braid fan*, denoted by Br_d, is the complete fan in $\mathbb{R}^{d+1}$ given by the hyperplanes $x_i - x_j = 0$ for all $i \neq j$.

Proposition 3.3.1 (Postnikov-Reiner-Williams)**.** *A polytope P in $V = \mathbb{R}^{d+1}$ is a generalized permutohedron if and only if its normal fan $\Sigma_V(P)$ with respect to V is refined by the Braid fan Br_d.*

Using the notation we give in Definition 3.2.8, the above result precisely says that the family of generalized permutohedra in $\mathbb{R}^{d+1}$ is $\mathrm{Poly}(\mathrm{Br}_d)$. Furthermore, it follows from [82, Proposition 2.6] that any usual permutohedron in $\mathbb{R}^{d+1}$ has the Braid fan

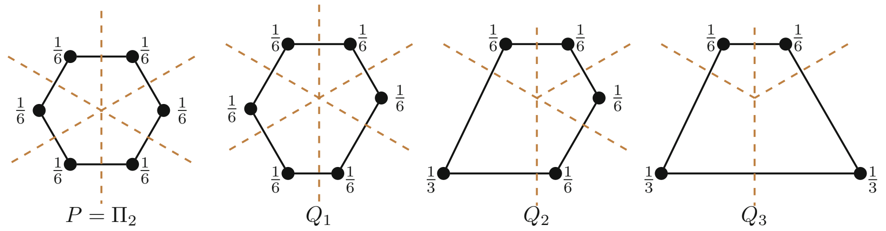


Fig. 3 Examples for Corollary 3.2.9

Br_d as its normal fan. In particular, the normal fan of the regular permutohedron Π_d is Br_d . In [24], Castillo and the author use these results together with the Reduction Theorem and its consequence (i.e., Corollary 3.2.9) to reduce Conjecture 3.1.1 to the following conjecture:

Conjecture 3.3.2 (Castillo-L.). *Every regular permutohedron Π_d is BV-α-positive.*

The following example demonstrates how Corollary 3.2.9 works and why Conjecture 3.1.1 can be reduced to Conjecture 3.3.2.

Example 3.3.3. Let P, Q_1, Q_2, and Q_3 be the 2-dimensional polytopes together with their normal fans shown in Fig. 3. One notices that P is the regular permutohedron Π_2 whose normal fan is Br_2, and each Q_i is a generalized permutohedron whose normal fan coarsens Br_2.

All the BV-α-values of the six vertices of P are $1/6$. Since Q_1 has the same normal fan as P, all of its six vertices also have the same BV-α-values. Now the normal fan of Q_2 coarsens that of P. In particular, if we let v be the vertex on the bottom-left of Q_2, then the normal cone $\mathrm{ncone}(v, Q_2)$ of Q_2 at v is the union of the normal cones of P at two of its vertices. It is a consequence of the "valuation properties" (P1) and (P2) that the BV-α-values $\alpha(v, Q_2)$ is the sum of the BV-α-values of these two vertices of P. Therefore, as shown in the figure, $\alpha(v, Q_2) = 1/6 + 1/6 = 1/3$. One sees that similar phenomenon happens for the polytope Q_3.

The above discussion shows that even if we did not know the BV-α-values of vertices of P, because each BV-α-value arising from Q_i is a summation of a subset of BV-α-values of vertices of P, BV-α-positivity of vertices of the regular permutohedron $P = \Pi_2$ would imply BV-α-positivity of vertices of the generalized permutohedron Q_i and thus would imply the constant Ehrhart coefficient of Q_i is positive.

Conjecture 3.3.2 was the main conjecture studied in [24], and partial progress was made on proving it, which gave us corresponding partial results on Conjecture 3.1.1.

3.3.2 Partial Results

The first approach of attacking Conjecture 3.3.2 is to directly compute BV-α-valuations. In order to do that, we need to compute the BV-construction Ψ. One

obvious benefit of considering Conjecture 3.3.2 instead of Conjecture 3.1.1 is that in each dimension there is only one regular permutohedron, and thus there are a limited number of BV-α-values or Ψ-values to be computed, especially for small d. Therefore, by explicit computation, we obtain the following result.

Theorem 3.3.4 (Castillo-L.). *For $d \leq 6$, the regular permutohedron Π_d is BV-α-positive. Therefore, all the integral generalized permutohedra (including matroid base polytopes) of dimension at most 6 are Ehrhart positive.*

Next, instead of focusing on all the coefficients of Ehrhart polynomials, we study certain special coefficients. Note that the first, second, and last Ehrhart coefficients are always positive, so we only consider other Ehrhart coefficients. Correspondingly, we need to know how to compute the BV-construction $\Psi(C)$ of cones C of dimension $2, 3, \dots, d-1$. The computation of the function Ψ is carried out recursively. Hence, it is quicker to compute Ψ for lower-dimensional cones. As a result, the value of $\alpha(F, P)$ is easier to compute if F is a higher-dimensional face.

In general, the computation of $\Psi(C)$ is quite complicated. However, when C is a unimodular cone, computations are greatly simplified. In dimensions 2 and 3, with the help of Maple code provided by Berline and Vergne, simple closed expression for Ψ of unimodular cones can be obtained [24, Lemmas 3.9 and 3.10]. Applying these formulas to Π_d, we obtain the second partial result toward Conjectures 3.3.2 and 3.1.1:

Theorem 3.3.5 (Castillo-L.). *For any d, and any face F of Π_d of codimension 2 or 3, we have $\alpha(F, \Pi_d)$ is positive, where α is the BV-α-valuation.*

Hence, the third and fourth Ehrhart coefficients of any integral generalized permutohedron (including matroid base polytopes) are positive.

Finally, the last partial result presented in [24] is the following:

Lemma 3.3.6 (Castillo-L.). *For any $d \leq 500$, and any edge E of Π_d, we have $\alpha(E, \Pi_d)$ is positive, where α is the BV-α-valuation.*

Hence, the linear Ehrhart coefficient of any integral generalized permutohedron (including matroid base polytopes) of dimension at most 500 is positive.

As we have discussed above, in order to compute the BV-α-values for an edge of a d-dimensional polytope, we have to compute the Ψ-value of a $(d-1)$-dimensional cone, which is extremely difficult for large d if we use Berline-Vergne algorithm directly. Therefore, we use a completely different strategy. Recall Property (P3) of the BV-construction, which says that Ψ is symmetric about coordinates. Note that the regular permutohedron Π_d is a polytope with much symmetry. So a lot of BV-α-values of Π_d coincide. In particular, we can separate edges of Π_d into to $\lceil \frac{d}{2} \rceil$ groups, where edges in each group share the same BV-α-values.

(Idea of Proof for Lemma 3.3.6). If we know the α-values for a give polytope P, Eq. (3.2) gives us a way to compute the Ehrhart coefficients. However, we can also use (3.2) in the other direction: Suppose we know the linear coefficient of $i(P, t)$, Eq. (3.2) gives us an equation for α-values arising from edges of P :

$$\sum_{E:\text{edge of } P} \alpha(E, P)\text{nvol}(E) = [t^1]i(P, t).$$

The α-values for the regular permutohedron also appear in other generalized permutohedra as all of them are in the family $\text{Poly}(\text{Br}_d)$. Therefore, if we can find $\lceil \frac{d}{2} \rceil$ "independent" generalized permutohedra for which we know their linear Ehrhart coefficients, then we can set up a $\lceil \frac{d}{2} \rceil \times \lceil \frac{d}{2} \rceil$ linear system for the $\lceil \frac{d}{2} \rceil$ α-values arising from edges of Π_d. Solving the system, we obtain all these α-values. See [24, Example 3.15] for an example of how we can solve a linear system to find α's.

Recall that Postnikov gives explicit formulas for the Ehrhart polynomials of type-$\mathcal{Y}$ generalized permutohedra (see Theorem 3.1.4). Among all the nontrivial Ehrhart coefficients, the linear terms can be easily described. Using these, we were able to set up, for each d, a desired linear system which is actually triangular. Solving the system for $d \leq 500$, we confirmed positivity of all $\lceil \frac{d}{2} \rceil$ α's arising from edges of Π_d. $\square$

Equivalence Statements. In addition to the partial results discussed above, two equivalent statements to Conjecture 3.3.2 were discovered. The first states that Conjecture 3.3.2 holds if and only if the mixed lattice point valuation on hypersimplices is positive [24, Corollary 5.6].

The second equivalent statement is in terms of Todd classes. The BV-construction gives one way to write the Todd class of the permutohedral variety in terms of the toric invariant cycles. We can show that if there is *any* way of writing such class as a positive combination of such cycles, then the BV-α-valuation is one of them. (See [25, Proposition 7.2] or [23].)

4 Negative Results

In this section, we will discuss examples and constructions of polytopes with negative Ehrhart coefficients. We start in Sect. 4.1 with the well-known Reeve tetrahedra, a family of 3-dimensional polytopes with negative linear Ehrhart coefficients. Constructions given in Sect. 4.2 were motivated by a refinement of Question 1.2, considering all possible sign patterns of Ehrhart coefficients. Examples studied in Sects. 4.3 and 4.4 provide negative answers to Question 1.2 for different families of polytopes (such as smooth polytopes and order polytopes), which will be summarized in Sect. 4.5. Finally in Sect. 4.6, we give negative examples addressing the question of whether Minkowski summation preserves Ehrhart positivity.

As mentioned before, due to the fact that the first, second, and last Ehrhart coefficients are always positive, given a d-dimensional polytope P, we need to ask the positivity question only for the coefficients of $t^{d-2}, t^{d-3}, \ldots, t^1$ in $i(P, t)$. We call these coefficients the *middle Ehrhart coefficients* of P.

4.1 Reeve Tetrahedra

For $d \leq 2$, there are no middle Ehrhart coefficients. Hence, possible examples with negative Ehrhart coefficients can appear only in dimension 3 or higher. The first example comes in dimension 3 : The *Reeve tetrahedron* T_m is the polytope with vertices $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, and $(1, 1, m)$, where m is a positive integer. Its Ehrhart polynomial is

$$i(T_m, t) = \frac{m}{6}t^3 + t^2 + \frac{12 - m}{6}t + 1.$$

One sees that the linear coefficient is 0 when $m = 12$ and is *negative* when $m \geq 13$.

4.2 Possible Sign Patterns

Motivated by the example of Reeve tetrahedra, Hibi, Higashitani, Tsuchiya, and Yoshida study possible sign patterns of middle Ehrhart coefficients and ask the following question:

Question 4.2.1 (Question 3.1 of [47]). Given a positive integer $d \geq 3$ and integers $1 \leq i_1 < \cdots < i_q \leq d - 2$, does there exist a d-dimensional integral polytope P such that the coefficients of $t^{i_1}, \dots, t^{i_q}$ of $i(P, t)$ are negative, and the remaining coefficients are positive?

The following is the main result in [47] providing a partial answer to Question 4.2.1.

Theorem 4.2.2 (Hibi-Higashitani-Tsuchiya-Yoshida). *Let $d \geq 3$. The following statements are true.*

(a) There exists an integral polytope P of dimension d such that all of its middle Ehrhart coefficients are negative.
(b) For each $1 \leq k \leq d - 2$, there exists an integral polytope P of dimension d such that $[t^k]i(P, t)$ is negative and all the remaining Ehrhart coefficients are positive.

The proof of both parts of the theorem is by construction. We will briefly discuss the construction for Theorem 4.2.2/(a) and refer interested readers to the original paper [47] for the other construction.

Sketch of proof for Theorem 4.2.2/(a). Let $L_n := [0, n]$, which is a 1-dimensional polytope, and its Ehrhart polynomial is $i(L_n, t) = nt + 1$. Define the polytope $P_m^{(d)}$ to be the direct product of $(d - 3)$ copies of L_{d-3} and one copy of the Reeve tetrahedron T_m. Then $P_m^{(d)}$ is a d-dimensional polytope with Ehrhart polynomial

$$i\left(P_m^{(d)}, t\right) = i(L_{d-3}, t)^{d-3} \cdot i(T_m, t) = ((d - 3)t + 1)^{d-3} \cdot \left(\frac{m}{6}t^3 + t^2 + \frac{12 - m}{6}t + 1\right).$$

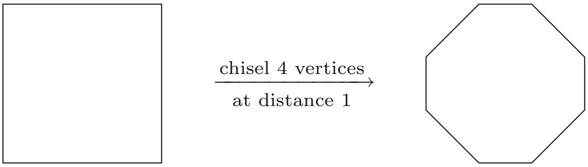


Fig. 4 From $3\square_2$ to $Q_2(3, 1)$

The coefficients of the Ehrhart polynomial of $P_m^{(d)}$ can be explicitly described, from which one can show that all middle Ehrhart coefficients are negative for sufficiently large m. $\square$

In addition to Theorem 4.2.2, Hibi et al. also show that answer to Question 4.2.1 is affirmative for $d \leq 6$ [47, Proposition 3.2]. Note that for $d \leq 6$, there are at most 3 middle Ehrhart coefficients. Later, Tsuchiya (private communication) improved their result showing that any sign pattern with at most 3 negatives is possible for the middle Ehrhart coefficients. Unfortunately, it is not currently clear how to extend the techniques used to prove this result to attack the question of whether any sign pattern with 4 negatives can occur. So Question 4.2.1 is still wide open.

4.3 Smooth Polytopes

A d-dimensional integral polytope P is called *smooth* (or *Delzant*) if each vertex is contained in precisely d edges, and the primitive edge directions form a lattice basis of $\mathbb{Z}^d$. In [18, Question 7.1], Bruns asked whether all smooth integral polytopes are Ehrhart positive. In [26], Castillo, Nill, Paffenholz, and the author show the answer is false by presenting counterexamples in dimensions 3 and higher. The main ideas we used was chiseling cubes and searching for negative BV-α-values.

The first set of examples we construct is as follows: For positive integers $a > 2b$, we let $Q_d(a, b)$ be the polytope obtained by chiseling *all* vertices of $a\square_d$ at distance b. (See Fig. 4 for an example.) Using inclusion-exclusion and the fact that the BV-α-values of cubes and standard simplices can be obtained easily due to property (P3), we obtain explicit formulas for all BV-α-values arising from $P_d(a, b)$, which we use to search for negative BV-α-values. The first negative values appear at $d = 7$, suggesting that we might have a negative Ehrhart coefficient in $Q_7(a, b)$. By direct computation, we are able to show that for some choices of (a, b), e.g., $(5, 2)$, the polytope $Q_d(a, b)$ has a negative linear Ehrhart coefficient for any $d \geq 7$. Therefore, we have the following result [26, Proposition 1.3]:

Proposition 4.3.1 (Castillo-L.-Nill-Paffenholz). *Let $\mathcal{N}_d$ be the normal fan of $Q_d(a, b)$. For $d \leq 6$, any d-dimensional smooth integral polytope with normal fan $\mathcal{N}_d$ is Ehrhart positive. For each $d \geq 7$, there exists a d-dimensional smooth integral polytope with normal fan $\mathcal{N}_d$ whose linear Ehrhart coefficient is negative.*

Remark 4.3.2. The polytope $Q_d(a, b)$ is not only a smooth polytope, but also a "type-B generalized permutohedron." The generalized permutohedra considered in Sect. 3 are of type A as the corresponding normal fan, Br_d, is constructed from the type A root system. As a consequence, a polytope P is a (type-A) generalized permutohedron if and only if each edge direction of P is of the form of $\boldsymbol{e}_i - \boldsymbol{e}_j$ for some $i \neq j$. Similarly, we can define a polytope P in $\mathbb{R}^d$ is a *type-B generalized permutohedron* if each edge direction of P is in the form of $\boldsymbol{e}_i \pm \boldsymbol{e}_j$ for some $i \neq j$ or of the form $\pm\boldsymbol{e}_i$ for some i. It is then straightforward to verify that $Q_d(a, b)$ is a type-B generalized permutohedron.

Using the idea of iterated chiseling cubes, we then improve the dimension range of our counterexamples from $d \geq 7$ to $d \geq 3$. (See [26, Section 2] for details.)

Theorem 4.3.3 (Castillo-L.-Nill-Paffenholz)**.** *For each $d \geq 3$, there exists a d-dimensional smooth integral polytope P such that all of its middle Ehrhart coefficients are negative.*

Note that the above theorem is a stronger version than part (a) of Theorem 4.2.2. Even though the original purpose of the paper [26] was to answer Bruns' question, in the process of searching for a counterexample, we obtained a separate result answering a different question. For positive integers $a > b$, we let $P_d(a, b)$ be the polytope obtained by chiseling *one* vertex off $a\square_d$ at distance b. It is clear that $P_d(a, b)$ share the same BV-α-values with $Q_d(a, b)$. Hence, it has negative BV-α-values at $d \geq 7$. However, it turns out any d-dimensional integral polytope P that has the same normal fan as $P_d(a, b)$ is Ehrhart positive [26, Lemma 3.9].

Corollary 4.3.4 (Castillo-L.-Nill-Paffenholz)**.** *For each $d \geq 7$, there exists a smooth projective fan Σ, such that its associated BV-α-values contains negative values, but any smooth integral polytope in* $\text{Poly}(\Sigma)$ *is Ehrhart positive.*

Therefore, BV-α-positivity is strictly stronger than Ehrhart positivity.

Finally, we studied a weaker version of Brun's question by requiring the smooth polytopes to be reflexive. More precisely, we asked whether all smooth reflexive polytopes have positive Ehrhart coefficients. Unfortunately, the answer to this question is still negative.

In fixed dimension d, there are only finitely many reflexive polytopes up to unimodular transformations. Because of their correspondence to toric Fano manifolds, smooth reflexive polytopes were completely classified up to dimension 9 [62, 76]. We used `polymake` [2] to verify that up to dimension 8 all of them are Ehrhart positive, but in dimension 9 the following counterexample came up [26]:

Example 4.3.5. Let P be the polytope in $\mathbb{R}^9$ defined by

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{pmatrix} \leq \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Using `polymake` [2], one can check that this polytope is smooth and reflexive with Ehrhart polynomial

$$\begin{aligned} i(P,t) = {} & 12477727/18144t^9 + 12477727/4032t^8 + 9074291/1512t^7 + 630095/96t^6 \\ & + 19058687/4320t^5 + 117857/64t^4 + 3838711/9072t^3 + 11915/1008t^2 \\ & - 6673/630t + 1, \end{aligned}$$

which has a negative linear coefficient.

4.4 Stanley's Example

In answering an Ehrhart positivity question posted on mathoverflow, Stanley gave the following example [100]:

Example 4.4.1. Let $\mathcal{Q}_k$ be the poset with one minimal element covered by k other elements. The Ehrhart polynomial of the order polytope $\mathcal{O}(\mathcal{Q}_k)$ is

$$i(\mathcal{O}(\mathcal{Q}_k), t) = \sum_{i=1}^{t+1} i^k.$$

One can compute that

$$[t^1]i(\mathcal{O}(\mathcal{Q}_{20}), t) = -168011/330 < 0.$$

Hence, the linear Ehrhart coefficient of $\mathcal{O}(\mathcal{Q}_k)$ is negative when $k = 20$.

Based on Stanley's example, the author and Tsuchiya studied the Ehrhart positivity question on polytopes $\mathcal{O}(\mathcal{Q}_k)$ for every k and gave a complete description of

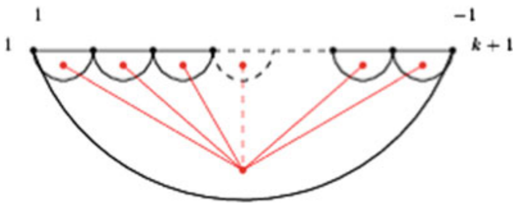


Fig. 5 $\mathcal{Q}_k$ and its corresponding planar graph G

which Ehrhart coefficients of $\mathcal{O}(\mathcal{Q}_k)$ are negative [61]. The following theorem is an immediate consequence to this description.

Theorem 4.4.2 (L.-Tsuchiya)**.** *The order polytope $\mathcal{O}(\mathcal{Q}_k)$ (defined in Example 4.4.1) is Ehrhart positive if and only if $k \leq 19$.*

Stanley's example and its extension are very interesting as $\mathcal{O}(\mathcal{Q}_k)$ belongs to a lot of different families of polytopes. First of all, it is an order polytope and thus is a (0, 1)-polytope.

Recall that a Gorenstein polytope of codegree s is an integral polytope such that sP is reflexive. It follows from a result by Hibi [42] that an order polytope is Gorenstein if and only if the underlying poset is *pure*; i.e., all maximal chains have the same length. Clearly, $\mathcal{Q}_k$ is pure. Thus, $\mathcal{O}(\mathcal{Q}_k)$ is a Gorenstein polytope.

Finally, Mészáros, Morales, and Striker proved a result observed by Postnikov establishing a connection between flow polytopes of planar graphs and order polytopes [72, Theorem 3.8]. Using this connection, Morales (private communication) observes that the order polytope $\mathcal{O}(\mathcal{Q}_k)$ is unimodularly equivalent to the flow polytope $\mathcal{F}_G(1, 0, \ldots, 0, -1)$, where G is the black graph on $k + 1$ vertices in Fig. 5. (Note the red part of the figure is $\mathcal{Q}_k$.)

4.5 Non-Ehrhart-Positive Families

For each of the families listed below, it is *not* true that all the integral polytopes in the family are Ehrhart positive.

(i) Smooth polytopes.
(ii) Type- B generalized permutohedra.
(iii) (0,1)-polytopes.
(iv) Order polytopes.
(v) Chain polytopes.
(vi) Flow polytopes.
(vii) Gorenstein polytopes.
(viii) Reflexive polytopes.
(ix) Smooth reflexive polytopes.

Furthermore, non-Ehrhart-positive examples were constructed for family (i) for each dimension $d \geq 3$, for family (ii) for each dimension $d \geq 7$, and for families (iii), (iv), (v), (vi), (vii), and (viii) for each dimension $d \geq 21$.

Proof. The conclusion for (i) and (ii) follows from Theorem 4.3.3, Proposition 4.3.1, and Remark 4.3.2. Notice that the order polytope $\mathcal{O}(\mathcal{Q}_k)$ considered in Sect. 4.4 has dimension $k+1$. Then the conclusion for (iii), (iv), (vi), and (vii) follows directly from discussion in Sect. 4.4. Next, (v) follows from (iv) and Remark 2.2.6, and (viii) follows from (vii), the connection between Gorenstein polytopes and reflexive polytopes and the fact that Ehrhart positivity is invariant under dilating operations. Finally, (ix) follows from Example 4.3.5.

4.6 Minkowski Sums

Recall that in Sect. 3.1.2, we learned that the type-$\mathcal{Y}$ generalized permutohedra which are defined to be Minkowski sums of dilated standard simplices are Ehrhart positive. Noticing that standard simplices are Ehrhart positive (see Sect. 2.1.2), we asked the following question in the first version of this survey:

Is it true that if two integral polytopes P and Q are Ehrhart positive, then their Minkowski sum $P+Q$ is Ehrhart positive?

Tsuchiya (private communication) constructed a few examples, which gave a negative answer to the above question, shortly after it was posted. Below are two of his examples.

Example 4.6.1. Let P be the 3-dimensional simplex with vertices

$$(0,0,0,0),\ (1,0,0,0),\ (0,1,0,0),\ (0,0,1,0),$$

and Q the 1-dimensional polytope with vertices

$$(0,0,0,0),\ (1,19,19,20).$$

It is easy to see that both P and Q are Ehrhart positive. However, one can check that $P+Q$ is a 4-dimensional polytope with Ehrhart polynomial

$$i(P+Q,t) = 10/3t^4 + 7/6t^3 - 1/3t^2 + 17/6t + 1,$$

which has a negative quadratic coefficient.

Example 4.6.2. Let P be the 5-dimensional simplex with vertices

$$(0,0,0,0,0),\ (1,0,0,0,0),\ (0,1,5,15,16),\ (0,0,1,0,0),\ (0,0,0,1,0),\ (0,0,0,0,1),$$

and Q the 5-dimensional simplex with vertices

$$(0,0,0,0,0),\ (1,0,0,0,0),\ (1,1,15,15,16),\ (0,0,1,0,0),\ (0,1,0,0,0),\ (0,0,0,0,1).$$

Since P is unimodularly equivalent to the standard 5-simplex, it is Ehrhart positive. Moreover, the Ehrhart polynomial of Q is

$$i(Q,t) = 1/8t^5 + 5/12t^4 + 17/24t^3 + 19/12t^2 + 13/6t + 1,$$

which also has positive coefficients.

However, $P + Q$ is a 5-dimensional polytope with Ehrhart polynomial

$$i(P+Q,t) = 3007/40t^5 + 359/24t^4 - 255/24t^3 + 193/24t^2 + 89/20t + 1,$$

which has a negative coefficient.

Notice that the polytopes given in Example 4.6.1 satisfy $\dim(P) + \dim(Q) = \dim(P+Q)$, and those in Example 4.6.2 satisfy $\dim(P) = \dim(Q) = \dim(P+Q)$. These are the two extreme situations in terms of dimensions. Therefore, even with some restrictions on the dimensions of P, Q, and $P+Q$, the answer to the question above is false.

5 Further Discussion

5.1 Ehrhart Positivity Conjectures

We list several families of polytopes that are conjectured to be Ehrhart positive.

5.1.1 Base-*r* Simplices

Recall the definition of $\Delta_{(1,\boldsymbol{q})}$ given in Sect. 2.2.2. For any positive integer $r \in \mathbb{P}$, we let $\boldsymbol{q}_r := (r-1, (r-1)r, (r-1)r^2, \ldots, (r-1)r^{d-1}) \in \mathbb{P}^d$ and then define the *base-r d-simplex* to be

$$\mathcal{B}_{(r,d)} := \Delta_{(1,\boldsymbol{q}_r)}.$$

Note that if $r = 1$, we obtain a polytope that is unimodularly equivalent to the standard d-simplex. The family of base-r d-simplices are introduced by Solus in his study of simplices for numeral systems [91], in which he shows that the h^*-polynomial of $\mathcal{B}_{(r,d)}$ is real-rooted and thus is unimodal. Based on computational evidence, Solus makes the following conjecture [91, Section 5]:

Conjecture 5.1.1 (Solus)**.** *The base-r d-simplex is Ehrhart positive.*

We remark that this family of $\Delta_{(1,\boldsymbol{q})}$ is very different from the ones constructed by Payne discussed in Example 2.2.5. If $r > 1$, the base-r simplex $\mathcal{B}_{(r,d)}$ always contains the origin as an interior point, and it follows from (1.2) that the degree of

h^*-polynomial of $\mathcal{B}_{(r,d)}$ is d. Since $\mathcal{B}_{(r,d)}$ is not reflexive, by Corollary 2.2.9 the roots of its h^*-polynomial are not all on the unit circle in the complex plane. Therefore, the techniques used to prove Ehrhart positivity for Payne's construction would not work here.

5.1.2 Birkhoff Polytopes

The *Birkhoff polytope* B_n is the convex polytope of $n \times n$ doubly stochastic matrices; that is, the set of real nonnegative matrices with all row and column sums equal to one. Equivalently, B_n can also be defined as the convex hull of all $n \times n$ permutation matrices. (See [106, Chaps. 5 and 6] for a detailed introduction to B_n.) There has been a lot of research on computing the volumes and Ehrhart polynomials of Birkhoff polytopes [8, 21, 32, 77]. The following conjecture was made by Stanley in a talk [97]:

Conjecture 5.1.2 (Stanley)**.** Birkhoff polytopes are Ehrhart positive.

By checking the available data [8], the first nine $i(B_n, t)$ have the property that all the roots have negative real parts. More importantly, Fig. 6 in [9] suggests that the roots of $i(B_n, t)$ form a certain pattern. Hence, it could be promising to use Lemma 2.2.1 to attack this conjecture.

We also remark that B_n is a Gorenstein polytope (up to lattice translation) of codegree n. However, with aforementioned data, one can see that B_n is not h^*-unit-circle-rooted. Hence, we cannot apply Theorem 2.2.2 to show that all roots of $i(B_n, t)$ have negative real parts.

5.1.3 Tesler Polytopes

For any $n \times n$ upper triangular matrix $M = (m_{i,j})$, the kth *hook sum* of M is the sum of all the elements on the kth row minus the sum of all the elements on the kth column excluding the diagonal entry:

$$(m_{k,k} + m_{k,k+1} + \cdots + m_{k,n}) - (m_{1,k} + m_{2,k} + \cdots + m_{k-1,k}).$$

For each $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{N}^n$, Mészáros, Morales, and Rhoades [71] define the *Tesler polytope*, denoted by $\mathrm{Tes}_n(\boldsymbol{a})$, to be the set of all $n \times n$ upper triangular matrices M with nonnegative entries and of *hook sum* $\boldsymbol{a}$; i.e., the kth hook sum of M is a_k. The lattice points in $\mathrm{Tes}_n(\boldsymbol{a})$ are exactly *Tesler matrices of hook sum* $\boldsymbol{a}$. When $\boldsymbol{a} = (1, 1, \ldots, 1)$, these are important objects in Haglund's work on diagonal harmonics [40]. Therefore, we call $\mathrm{Tes}_n(1, 1, \ldots, 1)$ the *Tesler matrix polytope* as Tesler matrices of hook sum $(1, 1, \ldots, 1)$ were the original Tesler matrices defined by Haglund.

Another interesting example of a Tesler polytope is $\mathrm{Tes}_n(1, 0, \ldots, 0)$, which turns out to be the *Chan-Robbins-Yuen* polytope or *CRY* polytope, a face of the Birkhoff

polytope. The *CRY* polytope, denoted by $\mathcal{CRY}_n$, is the convex hull of all the $n \times n$ permutation matrices $M = (m_{i,j})$ such that $m_{i,j} = 0$ if $i \geq j + 2$; i.e., all entries below the subdiagonal are zeros. It was initially introduced by Chan, Robbins, and Yuen in [28], in which they made an intriguing conjecture on a formula for the volume of $\mathcal{CRY}_n$ as a product of Catalan numbers. It was since proved by Zeilberger [107], Baldoni-Vergne [3], and Mészáros [68, 69].

Using the Maple code provided by Baldoni, Beck, Cochet, and Vergne [4], Morales computed the Ehrhart polynomials of both CRY polytopes and Tesler matrix polytopes for small n and made the following conjecture [73].

Conjecture 5.1.3 (Morales)**.** *For each positive integer n, the CRY polytope $\mathcal{CRY}_n =$ $\mathrm{Tes}_n(1, 0, \ldots, 0)$ and the Tesler matrix polytope $\mathrm{Tes}_n(1, 1, \ldots, 1)$ are both Ehrhart positive.*

Connection to flow polytopes. Mészáros et al. show in [71, Lemma 1.2] that for any $\boldsymbol{a} \in \mathbb{N}^n$, the Tesler polytope $\mathrm{Tes}_n(\boldsymbol{a})$ is unimodularly equivalent to the flow polytope $\mathcal{F}_{K_{n+1}}(\bar{\boldsymbol{a}})$, where K_{n+1} is the complete graph on $[n+1]$ and $\bar{\boldsymbol{a}}$ is defined as in (2.2). Therefore, Tesler polytopes are flow polytopes associated to complete graphs. Note that the complete graph does not satisfy the hypothesis of Corollary 2.1.5. So Conjecture 5.1.3 does not follow.

5.1.4 Stretched Littlewood-Richardson Coefficients

The Schur functions s_λ form a basis for the ring of symmetric functions. (See [96, Chap. 7] for background on symmetric functions.) Therefore, the product of two Schur functions s_λ and $\boldsymbol{s}_\mu$ can be uniquely expressed as

$$s_\lambda \cdot s_\mu = \sum_{\nu : |\nu| = |\lambda| + |\mu|} c^\nu_{\lambda,\mu} s_\nu .$$

We call the coefficients $c^\nu_{\lambda,\mu}$ in the above expression the *Littlewood-Richardson coefficients* or *LR coefficients*. There are many different ways of computing $c^\nu_{\lambda,\mu}$. For example, it counts the number of semistandard Young tableaux T of shape ν/λ with content μ such that the reading word of T satisfies the "Yamanouchi word condition" [65]. One immediate consequence of these descriptions is that the LR coefficients are nonnegative integers. In this chapter, we use the hive model [20, 53, 54] to describe the LR coefficients.

A *hive of size n* is a triangular array of numbers $a_{i,j}$ with $0 \leq i, j, i + j \leq n$ arranged on a triangular grid consisting of n^2 small equilateral triangles. See the left side of Fig. 6 for how a hive of size 4 should look like. For any adjacent triangles $\{a, b, c\}$ and $\{b, c, d\}$ in the hive, they form a rhombus $\{a, b, c, d\}$. The *hive condition* for this rhombus is

$$b + c \geq a + d. \tag{HC}$$

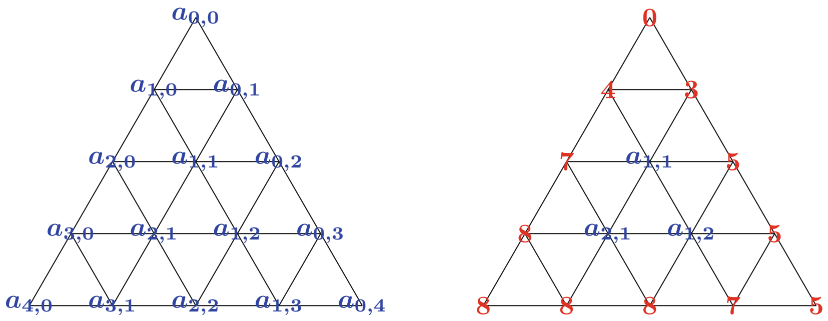


Fig. 6 A hive of size 4

Suppose $|\nu| = |\lambda| + |\mu|$ with $l(\nu), l(\lambda), l(\mu) \leq n$. A *Littlewood-Richardson-hive* or *LR-hive* of *type* (ν, λ, μ) is a hive $\{a_{i,j} \in \mathbb{N} : 0 \leq i, j, i + j \leq n\}$ with nonnegative integer entries satisfying the hive condition (HC) for all of its rhombi, with border entries determined by partitions ν, λ, μ in the following way: $a_{0,0} = 0$ and for each $j = 1, 2, \ldots, n$,

$$a_{j,0} - a_{j-1,0} = \nu_j, \quad a_{0,j} - a_{0,j-1} = \lambda_j, \quad a_{j,n-j} - a_{j-1,n-j+1} = \mu_k.$$

With this definition, the LR coefficient $c^{\nu}_{\lambda,\mu}$ counts the number of LR-hives of type (ν, λ, μ). (Note that this is independent from n as long as $l(\nu), l(\lambda), l(\mu) \leq n$.)

For example, if $\nu = (4, 3, 1)$, $\lambda = (3, 2)$ and $\mu = (2, 1)$, then the border of a corresponding LR-hive of size 4 is shown on the right side of Fig. 6. In fact, the hive condition will force $a_{2,1} = 8$ and $a_{1,2} = 7$. So it will be reduced to a hive of size 3. Finally, it follows from the hive condition that $6 \leq a_{1,1} \leq 7$. Thus, we have two LR-hives of this type, and we conclude that $c^{(4,3,1)}_{(3,2),(2,1)} = 2$.

From the above description, it is not hard to see that $c^{\nu}_{\lambda,\mu}$ counts the number of lattice points inside a polytope $P^{\nu}_{\lambda,\mu}$ determined by the border condition and the hive condition. Furthermore, for any positive integer t, the LR coefficient $c^{t\nu}_{t\lambda,t\mu}$ counts the number of lattice points inside the t^{th} dilation of $P^{\nu}_{\lambda,\mu}$:

$$c^{t\nu}_{t\lambda,t\mu} = |t P^{\nu}_{\lambda,\mu} \cap \mathbb{Z}^D|.$$

King, Tollu, and Toumazet studied $c^{t\nu}_{t\lambda,t\mu}$, which they call the *stretched Littlewood-Richardson coefficients* and made the following conjecture [51, Conjecture 3.1]:

Conjecture 5.1.4 (King-Tollu-Toumazet)**.** *For all partitions λ, μ, ν such that $c^{\nu}_{\lambda,\mu} > 0$, there exists a polynomial $f(t) = f^{\nu}_{\lambda,\mu}(t)$ in t such that $f(0) = 1$ and $f(t) = c^{t\nu}_{t\lambda,t\mu}$ for all positive integers t.*

Furthermore, all the coefficients of $f(t)$ are positive.

One notices that if $P^{\nu}_{\lambda,\mu}$ is an integral polytope, then the polynomiality part of the above conjecture follows from Ehrhart's theorem. However, in general, $P^{\nu}_{\lambda,\mu}$ is

a rational polytope, which only implies that $c_{t\lambda,t\mu}^{t\nu}$ is a quasi-polynomial with some period. Nevertheless, the assertion of polynomiality in the above conjecture was established first by Derksen and Weyman [33] using semi-invariants of quivers, and then by Rassart [86] using Steinberg's formula [104] and hyperplane arrangements. Hence, the polynomial asserted in Conjecture 5.1.4 can be considered to be an Ehrhart polynomial, and positivity assertion in the conjecture (which is still open) is exactly an Ehrhart positivity question.

5.2 Other Questions

Many questions related to Ehrhart positivity remain open. We include a few below.

5.2.1 Modified Bruns Question

As we have discussed in Sect. 4.3, the answer to Bruns' question of whether all smooth polytopes are Ehrhart positive is negative, where counterexamples are constructed for each dimension $d \geq 3$. Furthermore, we verify, with the help of `polymake` [2], that all smooth reflexive polytopes of dimension up to 8 are Ehrhart positive and that there exists a non-Ehrhart-positive smooth reflexive polytopes of dimension 9. However, we did not investigate smooth reflexive polytopes of higher dimensions. Therefore, one can ask:

Question 5.2.1. Does there exist a smooth reflexive polytope of dimension d with negative Ehrhart coefficients, for any $d \geq 10$?

Bruns' question can be rephrased using the language of fans: For any smooth projective fan Σ, is it true that *any* polytope with normal fan Σ is Ehrhart positive. Since the answer is false, a weaker version of this question can be asked:

Question 5.2.2. Is it true that for any smooth projective fan Σ, there exists *one* integral polytope P with normal fan Σ that is Ehrhart positive?

5.2.2 h^*-Vector for 3-Dimensional Polytopes

Here instead of studying Ehrhart positivity question for families of polytopes in which dimensions vary, we focus on polytopes with a fixed dimension and ask the following question:

Question 5.2.3. For each d, how likely is a d-dimensional integral polytope Ehrhart positive?

Since integral polytopes of dimension at most 2 are always Ehrhart positive, dimension 3 is a natural starting point.

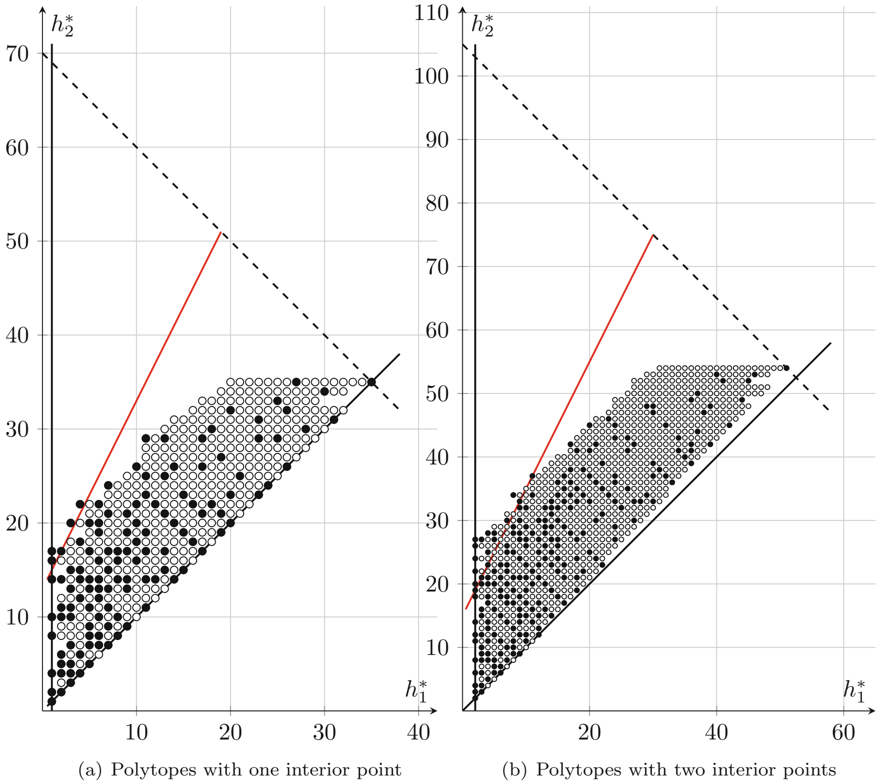


Fig. 7 The plot of (h_1^*, h_2^*) of 3-dimensional polytopes with 1 or 2 interior lattice points

We have mentioned in the introduction that various inequalities for h^*-vectors have been found. So we may use Formula (1.1) which gives a connection between the h^*-vector and Ehrhart coefficients together with known inequalities to study Question 5.2.3. Note that in dimension 3, only the linear Ehrhart coefficient could be negative. Applying (1.1), we obtain that P is Ehrhart positive (equivalently the linear Ehrhart coefficient of P is positive) if and only if the h^*-vector $(h_0^*, h_1^*, h_2^*, h_3^*)$ of P satisfies

$$11h_0^* + 2h_1^* - h_2^* + 2h_3^* > 0. \tag{5.1}$$

In [6], Balletti and Kasprzyk give classifications for 3-dimensional polytopes with 1 or 2 interior lattice points, using which they extract all possible h^*-vectors. Assume the number of interior lattice points is fixed to be 1 or 2. Applying (1.2), we obtain $h_0^* = 1$ and $h_3^* = 1$ or 2. Hence, only h_1^* and h_2^* change. Balletti and Kasprzyk then plot all occurring pairs of (h_1^*, h_2^*) in [6, Figure 5]. The black part of Fig. 7 is their figure, which we modify to include a red line representing the inequality (5.1), where points below the red line arise from polytopes with the Ehrhart positivity property.

Note that in each part of the figure, the big triangular area is bounded by three known inequalities for h^*-vectors. It is clear from the figure that these inequalities are far from optimal. Comparing the red line with the plotted data, one sees that a very high percentage of data points correspond to Ehrhart-positive polytopes. However, if we only look at the triangular region (without the data points), then the area below the red line has a much lower percentage of the region. Therefore, improving the inequality bounds for h^*-vectors will be helpful in understanding the Ehrhart positivity problem, in particular, in giving a more accurate answer to Question 5.2.3.

Acknowledgements The author is partially supported by a grant from the Simons Foundation #426756. The writing was completed when the author was attending the program "Geometric and Topological Combinatorics" at the Mathematical Sciences Research Institute in Berkeley, California, during the Fall 2017 semester, and she was partially supported by the NSF grant DMS-1440140.

The author would like to thank Gabriele Balletti, Ben Braun, Federico Castillo, Ron King, Akihiro Higashitani, Sam Hopkins, Thomas McConville, Karola Mészáros, Alejandro Morales, Benjamin Nill, Andreas Paffenholz, Alex Postnikov, Richard Stanley, Liam Solus, and Akiyoshi Tsuchiya for valuable discussions and helpful suggestions. The author is particularly grateful to Federico Castillo and Alejandro Morales for their help in putting together some figures and data used in this chapter. Finally, the author is thankful to the two anonymous referees for their careful reading of this chapter and various insightful comments and suggestions.

References

1. F. Ardila, C. Benedetti, J. Doker, Matroid polytopes and their volumes. Discret. Comput. Geom. **43**(4), 841–854 (2010). MR 2610473
2. B. Assarf, E. Gawrilow, K. Herr, M. Joswig, B. Lorenz, A. Paffenholz, T.Rehn, Polymake in linear and integer programming (2014), http://arxiv.org/abs/1408.4653
3. W. Baldoni, M. Vergne, Kostant partitions functions and flow polytopes. Transform. Groups **13**(3–4), 447–469 (2008). MR 2452600
4. M.W. Baldoni, M. Beck, C. Cochet, M. Vergne, Maple code for "volume computation for polytopes and partition functions for classical root systems". https://webusers.imj-prg.fr/~michele.vergne/IntegralPoints.html
5. G. Balletti, A. Higashitani, Universal inequalities in Ehrhart theory. arXiv:1703.09600
6. G. Balletti, A.M. Kasprzyk, Three-dimensional lattice polytopes with two interior lattice points. arXiv:1612.08918
7. M. Beck, Counting lattice points by means of the residue theorem. Ramanujan J. **4**(3), 299–310 (2000). MR 1797548
8. M. Beck, D. Pixton, The Ehrhart polynomial of the Birkhoff polytope. Discret. Comput. Geom. **30**(4), 623–637 (2003). MR (2013976)
9. M. Beck, J.A. De Loera, M. Develin, J. Pfeifle, R.P. Stanley, Coefficients and roots of Ehrhart polynomials. Contemp. Math. **374**, 15–36 (2005). MR 2134759
10. M. Beck, S. Robins, in *Computing the Continuous Discretely. Integer-Point Enumeration in Polyhedra*, 2nd edn. Undergraduate Texts in Mathematics (Springer, New York, 2015). With illustrations by David Austin. MR 3410115
11. N. Berline, M. Vergne, Local Euler-Maclaurin formula for polytopes. Mosc. Math. J. **7**(3), 355–386, 573 (2007). MR 2343137
12. C. Bey, M. Henk, J. Wills, Notes on the roots of Ehrhart polynomials. Discret. Comput. Geom. **38**(1), 81–98 (2007). MR 2322117

13. P. Brändén, in *Unimodality, Log-Concavity, Real-Rootedness and Beyond*. Handbook of Enumerative Combinatorics; Discrete Math. Appl. (Boca Raton) (CRC Press, Boca Raton, 2015), pp. 437–483. MR 3409348
14. B. Braun, Unimodality problems in Ehrhart theory, in *Recent Trends in Combinatorics*, vol. 159 (IMA Vol. Math. Appl.) (Sringer, Cham, 2016) pp. 687–711. MR 3526428
15. B. Braun, R. Davis, L. Solus, Detecting the integer decomposition property and Ehrhart unimodality in reflexive simplices, arXiv:1608.01614
16. F. Brenti, Log-concave and unimodal sequences in algebra, combinatorics, and geometry: an update, Jerusalem combinatorics, vol. 93. Contemporary Mathematics, vol. 178 (American Mathematical Society, Providence, 1994), pp. 71–89. MR 1310575
17. M. Brion, M. Vergne, An, equivariant Riemann-Roch theorem for complete, simplicial toric varieties. J. Reine Angew. Math. **482**, 67–92 (1997). MR 1427657
18. W. Bruns, The quest for counterexamples in toric geometry, Commutative algebra and algebraic geometry (CAAG-2010), Ramanujan Math. Soc. Lect. Notes Ser. **17**, 45–61 (2013) (Ramanujan Math. Soc., Mysore)
19. W. Bruns, J. Gubeladze, N.V. Trung, Normal polytopes, triangulations, and Koszul algebras. J. Reine Angew. Math. 485, 123–160 (1997), MR 1442191
20. A.S. Buch, The saturation conjecture (after A. Knutson and T. Tao). Enseign. Math. (2) **46**(1–2), 43–60 (2000) (With an appendix by William Fulton. MR 1769536)
21. E.R. Canfield, B.D. McKay, The asymptotic volume of the Birkhoff polytope. Online J. Anal. Comb. **4**, 4pp (2009). MR 2575172
22. S.E. Cappell, J.L. Shaneson, Genera of algebraic varieties and counting of lattice points. Bull. Am. Math. Soc. (N.S.) **30**(1), 62–69 (1994). MR 1217352
23. F. Castillo, Local Ehrhart Positivity, 91, Thesis (Ph.D.)-University of. California, Davis (2017) MR 3731970
24. F. Castillo, F. Liu, Berline-Vergne valuation and generalized permutohedra. Discret. Comput. Geom. **60**(4), 885–908 (2018)
25. F. Castillo, F. Liu, BV-α-positivity for generalized permutohedra. arXiv:1509.07884v1
26. F. Castillo, F. Liu, B. Nill, A. Paffenholz, Smooth polytopes with negative Ehrhart coeffcients. J. Comb. Theory Ser. A. **160**, 316–331(2018)
27. Castillo, F., & Liu, F. (2015). Ehrhart positivity for generalized permutohedra. *Discret. Math Theor. Comput. Sci. proc. FPSAC, '15*, 865–876.
28. C.S. Chan, D.P. Robbins, D.S. Yuen, On the volume of a certain polytope. Exp. Math. **9**(1), 91–99 (2000). MR 1758803
29. H. Conrads, Weighted, projective spaces and reflexive simplices. Manuscripta Math. **107**(2), pp. 215–227 (2002). MR 1894741
30. D.A. Cox, C. Haase, T. Hibi, A. Higashitani, Integer decomposition property of dilated polytopes. Electron. J. Comb. **21**(4), Paper 4.28, 17pp. (2014). MR 3292265,
31. J.A. De Loera, D.C. Haws, M. oeppe, Ehrhart polynomials of matroid polytopes and polymatroids. Discret. Comput. Geom. **42**(4), 670–702 (2009). MR 2556462
32. J.A. De Loera, F. Liu, R. Yoshida, A generating function for all semi-magic squares and the volume of the Birkhoff polytope. J. Algebr. Comb. **30**(1), 113–139 (2009). MR 2519852 (2010h:52017)
33. H. Derksen, J. Weyman, On the Littlewood-Richardson polynomials. J. Algebr. **255**(2), 247–257 (2002). MR 1935497
34. P. Diaconis, A. Gangolli, in *Rectangular Arrays with Fixed Margins. Discrete Probability and Algorithms* (Minneapolis, MN, 1993); IMA Volumes in Mathematics and its Applications, vol. 72 (Springer, New York, 1995), pp. 15–41. MR 1380519
35. Ehrhart, E. (1962). Sur les polyèdres rationnels homothétiques à n dimensions. *C. R. Acad. Sci. Paris*, *254*, 616–618.
36. S.E. Fienberg, U.E. Makov, M.M. Meyer, R.J. Steele, *Computing the Exact Distribution for a Multi-way Contingency Table Conditional on Its Marginal Totals*. Data Analysis from Statistical Foundations (Nova Science Publishers, Huntington, 2001), pp. 145–165. MR 2034512

37. W. Fulton, *Introduction to Toric Varieties, Annals of Mathematics Studies*, vol. 131 (Princeton University Press, Princeton, 1993) (The William H Roever Lectures in Geometry). MR 1234037
38. P. Galashin, S. Hopkins, T. McConville, A. Postnikov, Root system chip-firing I: interval-firing. arXiv:1708.04850
39. C. Haase, A. Paffenholz, L.C. Piechnik, F. Santos, Existence of unimodular triangulations—positive results, arXiv:1405.1687
40. J. Haglund, A polynomial expression for the Hilbert series of the quotient ring of diagonal coinvariants. Adv. Math. **227**(5), 2092–2106 (2011). MR 2803796
41. G. Hegedüs, A. Higashitani, A. Kasprzyk, Ehrhart polynomial roots of reflexive polytopes, arXiv:1503.05739
42. T. Hibi, Distributive lattices, affine semigroup rings and algebras with straightening laws. *Commutative Algebra And Combinatorics* (Kyoto, 1985). Advanced Studies in Pure Mathematics, vol. 11 (North-Holland, Amsterdam, 1987), pp. 93–109. MR 951198
43. T. Hibi, Some results on Ehrhart polynomials of convex polytopes. Discret. Math. **83**(1), 119–121 (1990). MR 1065691
44. T. Hibi, A lower bound theorem for Ehrhart polynomials of convex polytopes. Adv. Math. **105**(2), 162–165 (1994). MR 1275662
45. T. Hibi, Dual polytopes of rational convex polytopes. Combinatorica **12**(2), 237–240 (1992). MR 1179260
46. T. Hibi, Star-shaped complexes and Ehrhart polynomials. Proc. Am. Math. Soc. **123**(3), 723–726 (1995). MR 1249883
47. T. Hibi, A. Higashitani, A. Tsuchiya, K. Yoshida, Ehrhart polynomials with negative coefficients, arXiv:1506.00467
48. Hibi, T. (1992). *Algebraic Combinatorics of Convex Polytopes*. Australia: Carslaw Publications.
49. S. Hopkins, A. Postnikov, A positive formula for the Ehrhart-like polynomials from root system chip-firing, arXiv:1803.08472
50. J.M. Kantor, A. Khovanskii, Une application du théorème de Riemann-Roch combinatoire au polynôme d'Ehrhart des polytopes entiers de $\mathbf{R}^d$. C. R. Acad. Sci. Paris Sér. I Math. **317**(5), 501–507 (1993). MR 1239038
51. R.C. King, C. Tollu, F. Toumazet, Stretched Littlewood-Richardson and Kostka coefficients, symmetry in physics, in *CRM Proceedings and Lecture Notes*, vol. 34 (American Mathematical Society, Providence, 2004), pp. 99–112. MR 2056979
52. A.N. Kirillov, in *Ubiquity of Kostka Polynomials*, ed. by A.N. Kirillov, A. Tsuchiya, H. Umemura, Physics and Combinatorics, Proceedings Nagoya 1999. World Scientific (2001), arXiv:math.QA/9912094
53. A. Knutson, T. Tao, The honeycomb model of GLn(C) tensor products. I. Proof of the saturation conjecture. J. Am. Math. Soc. **12**(4), 1055–1090 (1999). MR 1671451
54. A. Knutson, T. Tao, C. Woodward, The honeycomb model of $\mathrm{GL}_n(C)$ tensor products. II. Puzzles determine facets of the Littlewood-Richardson cone. J. Am. Math. Soc. **17**(1), 19–48 (2004). MR 2015329
55. B.V. Lidskii, The Kostant function of the system of roots A_n. Funktsional. Anal. i Prilozhen. **18**(1), 76–77 (1984). MR 739099
56. F. Liu, A note on lattice-face polytopes and their Ehrhart polynomials. Proc. Am. Math. Soc. **137**(10), 3247–3258 (2009). MR 2515395 (2010h:52018)
57. F. Liu, Ehrhart polynomials of cyclic polytopes. J. Comb. Theory Ser. A **111**(1), 111–127 (2005). MR 2144858 (2006a:05012)
58. F. Liu, Ehrhart polynomials of lattice-face polytopes. Trans. Am. Math. Soc. **360**(6), 3041–3069 (2008). MR 2379786 (2009a:52012)
59. F. Liu, Higher integrality conditions, volumes and Ehrhart polynomials. Adv. Math. **226**(4), 3467–3494 (2011). MR 2764894 (2012a:52031)
60. F. Liu, L. Solus, On the relationship between Ehrhart unimodality and Ehrhart positivity. Annals of Combinatorics. arXiv:1804.08258

61. F. Liu, A. Tsuchiya, Stanley's non-Ehrhart-positive order polytopes, arXiv:1806.08403
62. B. Lorenz, A. Paffenholz, Smooth reflexive polytopes up to dimension 9 (2016), https://polymake.org/polytopes/paffenholz/www/fano.html
63. L. Lovász, M.D. Plummer, in *Matching Theory*. North-Holland Mathematics Studies, vol. 121 (North-Holland Publishing Co., Amsterdam; Akadémiai Kiadó) (Publishing House of the Hungarian Academy of Sciences), Budapest, 1986; Annals of Discrete Mathematics, 29). MR 859549
64. I.G. Macdonald, Polynomials associated with finite cell-complexes. J. Lond. Math. Soc. **4**(2), 181–192 (1971). MR 0298542
65. I.G. Macdonald, *Symmetric Functions and Hall Polynomials*, 2nd edn. Oxford Classic Texts in the Physical Sciences (The Clarendon Press, Oxford University Press, New York, 2015) (With contribution by A. V. Zelevinsky and a foreword by Richard Stanley, Reprint of the 2008 paperback edition [MR1354144]). MR 3443860
66. P. McMullen, The maximum numbers of faces of a convex polytope. Mathematika **17**, 179–184 (1970). MR 0283691
67. P. McMullen, *Valuations and Dissections*. Handbook of Convex Geometry, vol. A, B (North-Holland, Amsterdam, 1993), pp. 933–988. MR 1243000 (95f:52018)
68. K. Mészáros, On, product formulas for volumes of flow polytopes. *Combinatorial Methods in Topology and Algebra, Springer INdAM Series*, vol. 12 (Springer, Cham, 2015), pp. 91–95. MR 3467331
69. K. Mészáros, On, product formulas for volumes of flow polytopes. Proc. Am. Math. Soc. **143**(3), 937–954 (2015). MR 3293712
70. K. Mészáros, A.H. Morales, Volumes and Ehrhart polynomials of flow polytopes, arXiv:1710.00701
71. K. Mészáros, A.H. Morales, B. Rhoades, The polytope of Tesler matrices. Selecta Math. (N.S.) **23**(1), 425–454 (2017). MR 3595898
72. K. Mészáros, A.H. Morales, J. Striker, On flow polytopes, order polytopes, and certain faces of the alternating sign matrix polytope, arXiv:1510.03357
73. A.H. Morales, Ehrhart polynomials of examples of flow polytopes, https://sites.google.com/site/flowpolytopes/ehrhart
74. L. Moura, Polyhedral methods in design theory. Computational and Constructive Design Theory, Mathematics Applied, vol. 368 (Kluwer Academic Publishers, Dordrecht, 1996), pp. 227–254. MR 1398195
75. A. Nijehuis, H. Wilf, Representations of integers by linear forms in nonnegative integers. J Number Theory **4**, 98–106 (1972). MR 0288076
76. M. Øbro, An algorithm for the classification of smooth Fano polytopes, arXiv:0704.0049
77. I. Pak, Four questions on Birkhoff polytope. Ann. Comb. **4**(1), 83–90 (2000). MR 1763951
78. S. Payne, Ehrhart series and lattice triangulations. Discret. Comput. Geom. **40**(3), 365–376 (2008). MR 2443289
79. J. Pitman, R.P. Stanley, A polytope related to empirical distributions, plane trees, parking functions, and the associahedron. Discret. Comput. Geom. **27**(4), 603–634 (2002). MR 1902680 (2003e:52017)
80. J.E. Pommersheim, Toric varieties, lattice points and Dedekind sums. Math. Ann. **295**(1), 1–24 (1993). MR 1198839
81. J. Pommersheim, H. Thomas, Cycles representing the Todd class of a toric variety. J. Am. Math. Soc. **17**(4), 983–994 (2004). MR 2083474
82. A. Postnikov, Permutohedra, associahedra, and beyond. Int. Math. Res. Not. IMRN, **6**, 1026–1106 (2009). MR 2487491 (2010g:05399)
83. A. Postnikov, V. Reiner, L. Williams, Faces of generalized permutohedra. Doc. Math. **13**, 207–273 (2008). MR 2520477 (2010j:05425)
84. A. Postnikov, R.P. Stanley, Acyclic flow polytopes and Kostant's partition function. http://math.mit.edu/~rstan/transparencies/flow.ps
85. A.V. Pukhlikov, A.G. Khovanskii, The Riemann-Roch theorem for integrals and sums of quasipolynomials on virtual polytopes. Algebra i Analiz **4**(4), 188–216 (1992). MR 1190788

86. E. Rassart, A polynomiality property for Littlewood-Richardson coefficients. J. Comb. Theory Ser. A **107**(2), 161–179 (2004). MR 2078884
87. M.H. Ring, A. Schürmann, Local formulas for Ehrhart coefficients from lattice tiles, arXiv:1709.10390
88. Rodriguez-Villegas, F. R. (2002). On the zeros of certain polynomials. *Proc. Am. Math. Soc.*, *130*, 2251–2254.
89. J.R. Schmidt, A.M. Bincer, The Kostant partition function for simple Lie algebras. J. Math. Phys. **25**(8), 2367–2373 (1984). MR 751517
90. P.R. Scott, On, convex lattice polygons. Bull. Aust. Math. Soc. **15**(3), 395–399 (1976). MR 0430960
91. L. Solus, Simplices for numeral systems. Trans. Am. Math. Soc. arXiv:1706.00480 (to appear)
92. R.P. Stanley, A zonotope associated with graphical degree sequences, in *Applied Geometry and Discrete Mathematics*. DIMACS Series Discrete Mathematics & Theoretical Computer Science, vol. 4 (American Mathematical Society, Providence, 1991), pp. 555–570. MR 1116376 (92k:52020)
93. R.P. Stanley, Decompositions of rational convex polytopes. Ann. Discret. Math. **6**, 333–342 (1980). Combinatorial mathematics, optimal designs and their applications (Proceedings of the Symposium Combinatorial Mathematics and Optimal Design, Colorado State University, Fort Collins, Colorado, 1978). MR 593545
94. R.P. Stanley, *Enumerative Combinatorics*, vol. 1, 2nd edn. Cambridge Studies in Advanced Mathematics, vol. 49 (Cambridge University Press, Cambridge, 2012). MR 2868112
95. R.P. Stanley, Hilbert functions of graded algebras. Adv. Math. **28**(1), 57–83 (1978). MR 0485835
96. R.P. Stanley, in *Enumerative Combinatorics*, vol. 2. Cambridge Studies in Advanced Mathematics, vol. 62 (Cambridge University Press, Cambridge, 1999) (With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin). MR 1676282
97. R.P. Stanley, Lectures on lattice points in polytopes (2010), http://math.mit.edu/~rstan/transparencies/ehrhart1.pdf (2010)
98. R.P. Stanley, Log-concave and unimodal sequences in algebra, combinatorics, and geometry. Graph theory and its applications: East and West (Jinan, 1986). Ann. N Y Acad. Sci. **576**, 500–535 (1989). MR 1110850
99. R.P. Stanley, On the Hilbert function of a graded Cohen-Macaulay domain. J. Pure Appl. Algebr. **73**(3), 307–314 (1991). MR 1124790
100. R.P. Stanley, Positivity of Ehrhart polynomial coefficients. MathOverflow, https://mathoverflow.net/q/200574(version:2015-03-20)
101. R.P. Stanley, Two enumerative results on cycles of permutations. Eur. J. Comb. **32**(6), 937–943 (2011). MR 2821562
102. R.P. Stanley, Two poset polytopes. Discret. Comput. Geom. **1**(1), 9–23 (1986). MR 824105
103. A. Stapledon, Inequalities and Ehrhart δ-vectors. Trans. Am. Math. Soc. **361**(10), 5615–5626 (2009). MR 2515826
104. R. Steinberg, A general Clebsch-Gordan theorem. Bull. Am. Math. Soc. **67**, 406–407 (1961). MR 0126508
105. J. Treutlein, Lattice polytopes of degree 2. J. Comb. Theory Ser. A **117**(3), 354–360 (2010). MR 2592905
106. V.A. Yemelichev, M.M. Kovalëv, M.K. Kravtsov, in *Polytopes, Graphs and Optimisation* (Cambridge University Press, Cambridge, 1984), Translated from the Russian by G. H. Lawden, MR 744197
107. D. Zeilberger, Proof of a conjecture of Chan, Robbins, and Yuen. Electron. Trans. Numer. Anal. **9**, 147–148 (1999). Orthogonal polynomials: numerical and symbolic algorithms (Leganés, 1998). MR 1749805
108. G.M. Ziegler, *Lectures on Polytopes*. Graduate Texts in Mathematics, vol. 152 (Springer, New York, 1995). MR 1311028

Recent Trends in Quasisymmetric Functions

Sarah K. Mason

Abstract This article serves as an introduction to several recent developments in the study of quasisymmetric functions. The focus of this survey is on connections between quasisymmetric functions and the combinatorial Hopf algebra of noncommutative symmetric functions, appearances of quasisymmetric functions within the theory of Macdonald polynomials, and analogues of symmetric functions. Topics include the significance of quasisymmetric functions in representation theory (such as representations of the 0-Hecke algebra), recently discovered bases (including analogues of well-studied symmetric function bases), and applications to open problems in symmetric function theory.

1 Introduction

Quasisymmetric functions first appeared in the work of Stanley [117] and were formally developed in Gessel's seminal article on multipartite P-partitions [44]. Since their introduction, their prominence in the field of algebraic combinatorics has continued to grow. The number of recent developments in the study of quasisymmetric functions is far greater than would be reasonable to contain in this brief article; because of this, we choose to focus on a selection of subtopics within the theory of quasisymmetric functions. This article is skewed toward bases for quasisymmetric functions which are closely connected to Macdonald polynomials and the combinatorial Hopf algebra of noncommutative symmetric functions. A number of very interesting subtopics are therefore excluded from this article, including Stembridge's subalgebra of peak quasisymmetric functions [122] and its associated structure [16, 17, 19, 20, 34, 81], Ehrenborg's flag quasisymmetric function of a partially ordered set [34], colored quasisymmetric functions [69, 70], and type B quasisymmetric functions [25, 68, 100, 101]. This article also does not have the scope to address connections to probability theory such as riffle shuffles [120], random walks on

S. K. Mason (✉)
Department of Mathematics, Wake Forest University, Winston-Salem, NC, USA
e-mail: sarahkmason@gmail.com

H. Barcelo et al. (eds.), *Recent Trends in Algebraic Combinatorics*, Association for Women in Mathematics Series 16,
https://doi.org/10.1007/978-3-030-05141-9_7

quasisymmetric functions [64], or a number of other fascinating topics. Hopefully this article will inspire the reader to learn more about quasisymmetric functions and explore these topics in greater depth.

This article begins with an overview of symmetric functions. There are a number of excellent introductions to the subject including Fulton [39], Macdonald [87], Sagan [108], and Stanley [119]. The remainder of Sect. 1 deals with the genesis of quasisymmetric functions and several important bases. Section 2 discusses the significance of quasisymmetric functions in algebra and representation theory, while Sect. 3 explores connections to Macdonald polynomials. A number of recently introduced bases for quasisymmetric functions are described in Sect. 4. Section 5 is devoted to interactions with symmetric functions.

1.1 Basic Definitions and Background on Symmetric Functions

Recall that a *permutation* of the set $[n] := \{1, 2, \ldots, n\}$ is a bijection from the set $\{1, 2, \ldots, n\}$ to itself. The group of all permutations of an n-element set is denoted $\mathfrak{S}_n$. Let $\pi = \pi_1\pi_2 \cdots \pi_n \in \mathfrak{S}_n$ denote a permutation written in one-line notation. If $\pi_i > \pi_{i+1}$, then i is a *descent* of π. If $\pi_i > \pi_j$ and $1 \leq i < j \leq n$, then the pair (i, j) is an *inversion* of π. The *sign* of a permutation π (denoted $(-1)^{\pi}$) is the number of inversions of π.

Let $\mathbb{C}[x_1, x_2, \ldots, x_n]$ be the polynomial ring over the complex numbers $\mathbb{C}$ on a finite set of variables $\{x_1, x_2, \ldots, x_n\}$. A permutation $\pi \in \mathfrak{S}_n$ acts naturally on $f(x_1, x_2, \ldots, x_n) \in \mathbb{C}[x_1, x_2, \ldots, x_n]$ by

$$\pi f(x_1, x_2, \ldots, x_n) = f(x_{\pi_1}, x_{\pi_2}, \ldots x_{\pi_n}).$$

Definition 1.1. *The ring of* symmetric functions in n variables *(often denoted by Λ_n or Sym$_n$) is the subring of $\mathbb{C}[x_1, x_2, \ldots, x_n]$ consisting of all polynomials invariant under the above action for all permutations in $\mathfrak{S}_n$.*

This notion can be further extended to the ring *Sym* of symmetric functions in infinitely many variables. A *symmetric function* $f \in \text{Sym}$ is a formal power series $f \in \mathbb{C}[[X]]$ (with infinitely many variables $X = \{x_1, x_2, \ldots\}$) such that $f(x_1, x_2, \ldots) = f(x_{\pi_1}, x_{\pi_2}, \ldots)$ for every permutation of the positive integers.

A *partition* $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ of a positive integer n is a weakly decreasing sequence of positive integers which sum to n. The elements λ_i of the sequence are called the *parts* and the number of parts is called the *length* of the partition (denoted $\ell(\lambda)$). We write $\lambda \vdash n$ (or $|\lambda| = n$) to denote "λ is a partition of n."

Each partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ of n can be visualized as a *Ferrers diagram*, which consists of n squares (typically called *cells*) arranged into left-justified rows so that the ith row from the bottom contains λ_i cells. (Note that we are using French notation so that we think of the cells as indexed by their position in the coordinate

$$S = \begin{array}{l} \boxed{}\boxed{} \\ \boxed{}\boxed{}\boxed{}\boxed{} \\ \boxed{}\boxed{}\boxed{}\boxed{}\boxed{} \end{array} \qquad P = \begin{array}{l} \boxed{3}\boxed{1} \\ \boxed{3}\boxed{5}\boxed{5}\boxed{2} \\ \boxed{4}\boxed{1}\boxed{2}\boxed{7}\boxed{9} \end{array}$$

Fig. 1 S is the Ferrers diagram for (5, 4, 2) and P is a filling of (5, 4, 2)

$$T = \begin{array}{l} \boxed{}\boxed{}\boxed{} \\ \boxed{}\boxed{} \\ \boxed{}\boxed{}\boxed{}\boxed{} \end{array} \qquad F = \begin{array}{l} \boxed{4}\boxed{7}\boxed{1} \\ \boxed{2}\boxed{5} \\ \boxed{7}\boxed{5}\boxed{5}\boxed{3} \end{array}$$

Fig. 2 T is the composition diagram for (4, 2, 3) and F is a filling of (4, 2, 3)

plane. This means the cell (i, j) is the cell in the ith column from the left and the jth row from the bottom. A Ferrers diagram in English notation places the rows so that the ith row from the top contains λ_i cells, aligning with matrix indexing.) An assignment of positive integer entries to each of the cells in the Ferrers diagram of shape λ is called a *filling* (see Fig. 1).

A *composition* $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_\ell)$ of a positive integer n is a sequence of positive integers which sum to n. (It is sometimes necessary to allow 0 to appear as a part; a *weak composition* of a positive integer n is a sequence of *nonnegative* integers which sum to n.) The elements α_i of the sequence are called the *parts* and the number of parts is called the *length* of the composition. Write $\alpha \models n$ to denote "α is a composition of n." The *reverse*, α^*, of a composition α is obtained by reversing the order of the entries of α so that the last entry of α is the first entry of α^*, the second to last entry of α is the second entry of α^*, and so on. For example, if $\alpha = (4, 1, 3, 3, 2)$, then $\alpha^* = (2, 3, 3, 1, 4)$.

Each composition $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_\ell)$ of n can be visualized as a *composition diagram*, which consists of n cells arranged into left-justified rows so that the ith row from the bottom contains α_i cells, again using French notation. (At times we will shift to English notation, but the reader may assume we are using French notation unless specified otherwise.) An assignment of positive integer entries to each of the cells in the composition diagram of shape α is called a *filling*. See Fig. 2 for an example of a composition diagram and a filling.

The *refinement order* is a useful partial ordering on compositions. We say $\alpha \prec \beta$ in the refinement ordering if β can be obtained from α by summing adjacent parts of α. For example, $(2, 4, 1, 1, 3) \prec (2, 5, 4)$ and $(2, 4, 1, 1, 3) \prec (6, 5)$ but $(2, 5, 4) \not\prec (6, 5)$ and $(6, 5) \not\prec (2, 5, 4)$.

Let x^λ be the monomial $x_1^{\lambda_1} x_2^{\lambda_2} \cdots x_\ell^{\lambda_\ell}$. For example, if $\lambda = (6, 4, 3, 3, 1)$, then $x^\lambda = x_1^6 x_2^4 x_3^3 x_4^3 x_5$. One way to construct a symmetric function is to symmetrize such a monomial. The *monomial symmetric function* indexed by λ is

$$m_\lambda(X) = \sum x_{i_1}^{\lambda_1} x_{i_2}^{\lambda_2} \cdots x_{i_\ell}^{\lambda_\ell},$$

where the sum is over all distinct monomials with exponents $\lambda_1, \lambda_2, \dots, \lambda_\ell$.

The monomial symmetric functions $\{m_\lambda \mid \lambda \vdash n\}$ form a basis for the vector space Sym^n of degree n symmetric functions. There are many elegant and useful bases for symmetric functions including three multiplicative bases obtained by describing the

basis element f_n and then setting $f_\lambda = f_{\lambda_1} f_{\lambda_2} \cdots f_{\lambda_\ell}$. For example, the *elementary symmetric functions* $\{e_\lambda \mid \lambda \vdash n\}$ are defined by setting $e_n = m_{(1^n)}$, the *complete homogeneous symmetric functions* $\{h_\lambda \mid \lambda \vdash n\}$ are defined by setting $h_n = \sum_{\lambda \vdash n} m_\lambda$, and the *power sum symmetric functions* $\{p_\lambda \mid \lambda \vdash n\}$ are obtained by setting $p_n = m_{(n)}$. Notice that the monomial symmetric functions are not multiplicative; that is, m_λ is not necessarily equal to $m_{\lambda_1} m_{\lambda_2} \cdots m_{\lambda_\ell}$.

Define a scalar product (a bilinear form $\langle f, g\rangle$ with values in $\mathbb{Q}$, sometimes referred to as an *inner product*) on Sym by requiring that

$$\langle h_\lambda, m_\mu \rangle = \delta_{\lambda\mu},$$

where δ is the Kronecker delta. This means the complete homogeneous and monomial symmetric functions are dual to each other under this scalar product. The power sums are orthogonal under this scalar product. This means $\langle p_\lambda, p_\mu \rangle = \delta_{\lambda\mu} z_\lambda$, where $z_\lambda = 1^{m_1}(m_1!) 2^{m_2}(m_2)! \cdots k^{m_k}(m_k!)$ with m_i equal to the number of times the value i appears in λ. For example,

$$z_{(4,4,4,2,1,1)} = 1^2 2! 2^1 1! 3^0 0! 4^3 3! = 1536.$$

(Note that $\frac{n!}{z_\lambda}$ counts the number of permutations of cycle type λ [14, 108].)

Let $\omega\colon \text{Sym} \to \text{Sym}$ be the involution on symmetric functions defined by $\omega(e_n) = h_n$. (Note that this implies $\omega(e_\lambda) = h_\lambda$ for all partitions λ.) Then $\omega(p_n) = (-1)^{n-1} p_n$.

1.2 Schur Functions

The *Schur function* basis is one of the most important bases for symmetric functions due to its deep connections to representation theory and geometry as well as its combinatorial properties. Schur functions are orthonormal under the scalar product described above and can be defined in a number of different ways. We begin with a combinatorial description, for which we will need several definitions.

A filling of a partition diagram λ in such a way that the row entries are weakly increasing from left to right and the column entries are strictly increasing from bottom to top is called a *semi-standard Young tableau* of shape λ (see Fig. 3). The *content* of such a filling is the composition $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, where α_i is the number of times the entry i appears in the filling. The *weight* of a semi-standard Young tableau T is the monomial $x^T = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k}$. The set of all semi-standard Young tableaux of shape λ is denoted $\mathrm{SSYT}(\lambda)$. A semi-standard Young tableau of shape λ in which each entry from 1 to n (where $n = |\lambda|$) appears exactly once is called a *standard Young tableau*, and the set of all standard Young tableaux of shape λ is denoted $\mathrm{SYT}(\lambda)$.

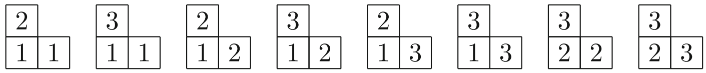


Fig. 3 Set of all semi-standard Young tableaux of shape (2, 1) whose entries are in the set $\{1, 2, 3\}$

Definition 1.2. *The* Schur function $s_\lambda(x_1, x_2, \ldots, x_n)$ *is the generating function for the weights of all semi-standard Young tableaux of shape* λ *with entries in the set* $\{1, 2, \ldots, n\}$*; that is*

$$s_\lambda(x_1, x_2, \ldots, x_n) = \sum_{T \in SSYT(\lambda)} x^T.$$

Here the sum is over all semi-standard Young tableaux whose entries are in the set $[n]$. *Extend this definition to infinitely many variables by allowing entries from the set of all positive integers.*

Figure 3 shows that

$$s_{21}(x_1, x_2, x_3) = x_1^2x_2 + x_1^2x_3 + x_1x_2^2 + 2x_1x_2x_3 + x_1x_3^2 + x_2^2x_3 + x_2x_3^2.$$

Notice that $s_{(1^n)} = m_{(1^n)} = e_n$, since $s_{(1^n)}$ is constructed by filling a vertical column with positive integers so that no entries repeat, and $s_n = h_n$ since s_n is constructed by filling a horizontal row with weakly increasing positive integers. The Schur functions enjoy a large number of beautiful properties, including the celebrated Littlewood–Richardson formula for the coefficients appearing in the product of two Schur functions (which can also be computed algorithmically using the Remmel–Whitney rule [106]) and the Robinson–Schensted–Knuth Algorithm [75, 110] which maps bijectively between matrices with finite nonnegative integer support and pairs (P, Q) of semi-standard Young tableaux of the same shape.

The Schur functions were classically described as quotients involving the Vandermonde determinant and can be defined in a number of other ways. One method of construction that can readily be generalized to other settings is through *Bernstein creation operators*.

Theorem 1.3. ([125]) *Define an operator* $\mathbf{B}_m : \mathrm{Sym}^n \to \mathrm{Sym}^{m+n}$ *by*

$$\mathbf{B}_m := \sum_{i \geq 0} (-1)^i h_{m+i} e_i^\perp,$$

where $f^\perp : \mathrm{Sym} \to \mathrm{Sym}$ *is an operator defined by* $\langle fg, h \rangle = \langle g, f^\perp h \rangle$ *for all* $g, h \in$ *Sym. Then for all tuples* $\alpha \in \mathbb{Z}^m$,

$$s_\alpha = \mathbf{B}_{\alpha_1} \mathbf{B}_{\alpha_2} \cdots \mathbf{B}_{\alpha_m}(1).$$

Note that this method for constructing Schur functions is more general than the combinatorial method described above, because Bernstein creation operators define Schur functions indexed by tuples of nonnegative integers (weak compositions) rather than just partitions.

Schur functions appear in many areas of mathematics beyond combinatorics. They correspond to characters of irreducible representations of the general linear group. Their multiplicative structure describes the cohomology of the Grassmannian of subspaces of a vector space. See the comprehensive texts by Fulton [39] and Sagan [108] for more details about Schur functions and their roles in combinatorics, representation theory, and geometry.

1.3 Quasisymmetric Functions

The ring Sym of symmetric functions is contained inside a larger ring of *quasisymmetric functions*, denoted by QSym, which can be thought of as all bounded degree formal power series f on an infinite alphabet $x_1, x_2, \dots$ such that the coefficient of $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k}$ in f is equal to coefficient of $x_{j_1}^{\alpha_1} x_{j_2}^{\alpha_2} \cdots x_{j_k}^{\alpha_k}$ in f for any sequence of positive integers $1 \leq j_1 < j_2 < \cdots < j_k$ and any composition $(\alpha_1, \alpha_2, \dots, \alpha_k)$. It is often convenient to restrict to n variables so that $f \in \mathrm{QSym}_n$ if and only if the coefficient of $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k}$ in f is equal to the coefficient of $x_{j_1}^{\alpha_1} x_{j_2}^{\alpha_2} \cdots x_{j_k}^{\alpha_k}$ in f for any sequence of positive integers $1 \leq j_1 < j_2 < \cdots < j_k \leq n$ and any composition $(\alpha_1, \alpha_2, \dots, \alpha_k)$.

For example, the polynomial

$$f(x_1, x_2, x_3) = x_1^3 x_2^5 + x_1^3 x_3^5 + x_2^3 x_3^5$$

is in QSym_3, as is

$$g(x_1, x_2, x_3) = x_1^3 x_2^5 + x_1^3 x_3^5 + x_2^3 x_3^5 + x_1^5 x_2^3 + x_1^5 x_3^3 + x_2^5 x_3^3,$$

but

$$h(x_1, x_2, x_3) = x_1^3 x_2^5 + x_1^3 x_3^5 + x_2^5 x_3^3$$

is not quasisymmetric in three variables since $x_2^3 x_3^5$ does not appear, and neither do $x_1^5 x_2^3$ and $x_1^5 x_3^3$. The quasisymmetric functions in n variables are precisely the functions invariant under a quasisymmetrizing action of the symmetric group $\mathfrak{S}_n$ introduced by Hivert [65, 66].

The origins of quasisymmetric functions first appeared in Stanley's work on P-partitions [117]. Gessel introduced the ring of quasisymmetric functions through his generating functions for Stanley's P-partitions [44]. (See Gessel [45] for a historical survey of P-partitions.) To be precise, let $[m]$ be the set $\{1, 2, \dots, m\}$ and let X be an infinite totally ordered set such as the positive integers. A *partially ordered set* (or *poset*), $(P, <_P)$, is a set of elements P and a partial ordering $\leq_P$ satisfying:

- reflexivity ($\forall x \in P, x \leq_P x$),
- antisymmetry (if $x \leq_P y$ and $y \leq_P x$, then $x = y$), and
- transitivity (if $x \leq_P y$ and $y \leq_P z$, then $x \leq_P z$).

Write $x <_P y$ if $x \leq_P y$ and $x \neq y$. A P-partition of a poset P with elements $[m]$ and partial order $\leq_P$ is a function $f : [m] \to X$ such that:

(1) $i <_P j$ implies $f(i)$ is less than or equal to $f(j)$, and
(1) $i <_P j$ and $i > j$ (under the usual ordering on integers) implies $f(i)$ is strictly less than $f(j)$.

Each permutation π corresponds to a totally ordered poset P_π where $\pi_1 <_\pi \pi_2 <_\pi \cdots <_\pi \pi_m$. It is these permutation posets that are used to construct Gessel's *fundamental quasisymmetric functions* F_α. To do this, give each P_π-partition f a *weight* $x^f = \prod x_{f(i)}$ and sum the weights over all P_π-partitions.

For example, let $\pi = 312$ (written in one-line notation). Condition (1) implies that $f(3) \leq f(1) \leq f(2)$. Condition (2) implies $f(3) < f(1)$. Therefore, $f(3) < f(1) \leq f(2)$, and the following table depicts the P_π-partitions involving the subset $\{1, 2, 3\}$ of X.

f(3)	f(1)	f(2)	x^f
1	2	2	$x_1x_2^2$
1	2	3	$x_1x_2x_3$
1	3	3	$x_1x_3^2$
2	3	3	$x_2x_3^2$

Therefore, the fundamental quasisymmetric function corresponding to $\pi = (3, 1, 2)$ and restricted to three variables is

$$x_1x_2^2 + x_1x_2x_3 + x_1x_3^2 + x_2x_3^2.$$

Note that this function depends only on the descent set of π and not on the underlying permutation; therefore, the indexing set for these generating functions is the set of all subsets of the set $[m-1]$ together with the number m to indicate the degrees of the monomials. We use an ordered pair consisting of a capital letter together with m when indexing by sets. In this paper, we use Greek letters to denote compositions, but in other places, such as [43], capital letters are used.

Definition 1.4. *([44]) Let L be a subset of* $[m-1]$. *Then*

$$F_{L,m} = \sum_{\substack{i_1 \leq i_2 \leq \cdots \leq i_m \\ i_j < i_{j+1} \text{ if } j \in L}} x_{i_1} x_{i_2} \cdots x_{i_m}$$

Every subset $L = \{L_1, L_2, \dots, L_k\}$ of the set $[m-1]$ corresponds to a composition $\beta(L) = (L_1, L_2 - L_1, \cdots, m - L_k)$, and therefore, the fundamental quasisymmetric functions are often indexed by compositions rather than sets. The fundamental quasisymmetric functions are homogeneous of degree m; the value m is apparent when the index is a composition α since $|\alpha| = m$. When the index is given by a subset, this value m must be specified. For example, if $m = 4$ then

$$F_{\{2,3\},4}(x_1, x_2, x_3, x_4) = \sum_{i_1 \le i_2 < i_3 < i_4} x_{i_1} x_{i_2} x_{i_3} x_{i_4}$$
$$= x_1^2 x_2 x_3 + x_1^2 x_2 x_4 + x_1^2 x_3 x_4 + x_1 x_2 x_3 x_4 + x_2^2 x_3 x_4.$$

whereas if $m = 5$ then

$$F_{\{2,3\},5}(x_1, x_2, x_3, x_4) = \sum_{i_1 \le i_2 < i_3 < i_4 \le i_5} x_{i_1} x_{i_2} x_{i_3} x_{i_4} x_{i_5}$$
$$= x_1^2 x_2 x_3^2 + x_1^2 x_2 x_3 x_4 + x_1^2 x_2 x_4^2 + x_1^2 x_3 x_4^2 + x_1 x_2 x_3 x_4^2 + x_2^2 x_3 x_4^2.$$

The *monomial quasisymmetric function*, M_α, in infinitely many variables $\{x_1, x_2, \dots\}$ and indexed by the composition $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$, is obtained by quasisymmetrizing the monomial $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k}$. That is,

$$M_\alpha(x_1, x_2, \dots) = \sum_{i_1 < i_2 < \cdots < i_k} x_{i_1}^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_k}^{\alpha_k}.$$

This definition can be restricted to finitely many variables by requiring that $i_k \le n$. Note that $m_\lambda = \sum_{\tilde{\alpha} = \lambda} M_\alpha$, where $\tilde{\alpha}$ is the partition obtained by rearranging the parts of α into weakly decreasing order. Every fundamental quasisymmetric function can be written as a positive sum of monomial quasisymmetric functions as follows:

$$F_\alpha = \sum_{\beta \preceq \alpha} M_\beta.$$

The Schur functions decompose into a positive sum of fundamental quasisymmetric functions; to describe this decomposition, we need one additional definition. Each standard Young tableau T has an associated *descent set* $D(T)$ given by $i \in D(T)$ if and only if $i + 1$ appears in a higher row of T than i. Then

$$s_\lambda = \sum_{T \in \mathrm{SYT}(\lambda)} F_{D(T), |\lambda|}. \tag{1.1}$$

For example, if $\lambda = (3, 2)$, then the standard Young tableaux of shape $(3, 2)$ are shown in Fig. 4, with respective descent sets $\{3\}$, $\{2, 4\}$, $\{2\}$, $\{1, 4\}$, and $\{1, 3\}$. Therefore,

$$\begin{array}{|c|c|c}\hline 4 & 5 & \multicolumn{1}{|c}{} \\ \hline 1 & 2 & \multicolumn{1}{|c|}{3} \\ \hline \end{array} \qquad \begin{array}{|c|c|c}\hline 3 & 5 & \multicolumn{1}{|c}{} \\ \hline 1 & 2 & \multicolumn{1}{|c|}{4} \\ \hline \end{array} \qquad \begin{array}{|c|c|c}\hline 3 & 4 & \multicolumn{1}{|c}{} \\ \hline 1 & 2 & \multicolumn{1}{|c|}{5} \\ \hline \end{array} \qquad \begin{array}{|c|c|c}\hline 2 & 5 & \multicolumn{1}{|c}{} \\ \hline 1 & 3 & \multicolumn{1}{|c|}{4} \\ \hline \end{array} \qquad \begin{array}{|c|c|c}\hline 2 & 4 & \multicolumn{1}{|c}{} \\ \hline 1 & 3 & \multicolumn{1}{|c|}{5} \\ \hline \end{array}$$

Fig. 4 Five standard Young tableau of shape $(3, 2)$

$$\begin{aligned} s_{3,2} &= F_{\{3\},5} + F_{\{2,4\},5} + F_{\{2\},5} + F_{\{1,4\},5} + F_{\{1,3\},5} \\ &= F_{32} + F_{221} + F_{23} + F_{131} + F_{122}. \end{aligned}$$

Valuable information can be gained about symmetric functions by examining their expansion into quasisymmetric functions, especially into the fundamental quasisymmetric functions. For example, a symmetric function is said to be *Schur-positive* if it can be written as a positive sum of Schur functions. Schur positivity is important because of its deep connection to representations of the symmetric group. Assaf's recently developed paradigm called *dual equivalence* [7] provides machinery to prove that a function is Schur-positive based on its expansion into the fundamentals and their connection to objects called dual equivalence graphs. The *Eulerian quasisymmetric functions* [113] are defined as sums of the fundamental quasisymmetric functions indexed by certain permutation statistics. Eulerian quasisymmetric functions are in fact always symmetric. Their generating functions are deeply connected to Euler's exponential generating functions for the Eulerian polynomials. Eulerian quasisymmetric functions can also be used to refine a number of classical results on permutation statistics. We will not be able to address these topics in this brief survey article but encourage the interested reader to see [7, 113] for details.

2 Algebra and Representation Theory

Even before they were formally defined, quasisymmetric functions appeared naturally in algebraic settings. The *Leibniz–Hopf algebra* is the free associative algebra over the integers which in fact is isomorphic to the algebra of noncommutative symmetric functions, which we shall define in Sect. 2.1. In 1972, Ditters claimed [30, Proposition 2.2] that the Leibniz–Hopf algebra is a free commutative algebra over the integers. This statement was later referred to as the *Ditters Conjecture* due to an error in the original proof and was then proved by Hazewinkel [60, 62] using combinatorial techniques and later by Baker and Richter [10] using methods from algebraic topology. Malvenuto and Reutenauer [90, Corollary 2.2] prove that QSym is a free module over Sym. In fact, Garsia and Wallach [41] further show that the quotient QSym_n over Sym_n has dimension $n!$. Aval and Bergeron [9] prove that the quotient of $\mathbb{Z}[x_1, x_2, \ldots, x_n]$ modulo quasisymmetric functions in n variables with no constant term has Hilbert series $\sum C_n t^n$ where C_n is the nth Catalan number (see also [8] for the case with infinitely many variables). An explicit basis for the quotient space is given in [79].

$$\begin{tikzcd} \mathcal{A}\otimes\mathcal{A}\otimes\mathcal{A} \arrow[r, "m\otimes id"] \arrow[d, "id\otimes m"] & \mathcal{A}\otimes\mathcal{A} \arrow[d, "m"] \\ \mathcal{A}\otimes\mathcal{A} \arrow[r, "m"] & \mathcal{A} \end{tikzcd}$$

(a) Associativity property

$$\begin{tikzcd} & \mathcal{A}\otimes\mathcal{A} \arrow[dd, "m"] & \\ R\otimes\mathcal{A} \arrow[ur, "u\otimes id"] \arrow[dr, "s"] & & \mathcal{A}\otimes R \arrow[ul, "id\otimes u"'] \arrow[dl, "s"] \\ & \mathcal{A} & \end{tikzcd}$$

(b) Unitary property

Fig. 5 Commutative diagrams for algebras

$$\begin{tikzcd} \mathcal{C}\otimes\mathcal{C}\otimes\mathcal{C} & \mathcal{C}\otimes\mathcal{C} \arrow[l, "\Delta\otimes id"] \\ \mathcal{C}\otimes\mathcal{C} \arrow[u, "id\otimes\Delta"] & \mathcal{C} \arrow[l, "\Delta"] \arrow[u, "\Delta"'] \end{tikzcd}$$

(a) Coassociativity property

$$\begin{tikzcd} & \mathcal{C}\otimes\mathcal{C} \arrow[dl, "\varepsilon\otimes id"'] \arrow[dr, "id\otimes\varepsilon"] & \\ R\otimes\mathcal{C} & & \mathcal{C}\otimes R \\ & \mathcal{C} \arrow[uu, "\Delta"] \arrow[ul, "1\otimes"] \arrow[ur, "\otimes 1"'] & \end{tikzcd}$$

(b) Counitary property

Fig. 6 Commutative diagrams for coalgebras (Here $\otimes 1$ applied to an element $c \in \mathcal{C}$ means $c \otimes 1$ and $1\otimes$ applied to an element $c \in \mathcal{C}$ means $1 \otimes c$.)

The ring of quasisymmetric functions (QSym) is an important example of a combinatorial Hopf algebra (discussed in Sect. 2.1). QSym is closely connected to Solomon's descent algebra (described in Sect. 2.2) and plays a role in representations of the 0-Hecke algebra (see Sect. 2.3).

2.1 Combinatorial Hopf Algebras

The following definitions, leading to the description of a combinatorial Hopf algebra, closely follow the expositions in [47, 85].

Let R be a commutative ring with an identity element. An *associative algebra* over R is an R-module $\mathcal{A}$ together with a *product* (or *multiplication*) $m\colon \mathcal{A}\otimes\mathcal{A} \to \mathcal{A}$ and a *unit* $u\colon\ R \to \mathcal{A}$ satisfying associativity ($m(m(a, b), c) = m(a, m(b, c))$) and a unitary property which implies that the unit map commutes with scalar multiplication. To be precise, m and u are R-linear maps such that if id is the identity map on $\mathcal{A}$ and s is scalar multiplication, then the diagrams in Fig. 5 commute.

A *coalgebra* over R is an R-module $\mathcal{C}$ together with a *coproduct* (or *comultiplication*) $\Delta\colon \mathcal{C} \to \mathcal{C}\otimes\mathcal{C}$ and a *counit* (or *augmentation*) $\varepsilon\colon \mathcal{C} \to R$ satisfying a *coassociativity* property and a *counitary* property so that when the directions of the arrows in Fig. 5 are reversed, the resulting diagrams (shown in Fig. 6) commute.

An *algebra morphism* is a map $f\colon \mathcal{A} \to \mathcal{A}'$ from an R-algebra $(\mathcal{A}, m, u)$ to another R-algebra $(\mathcal{A}', m', u')$ such that

$$f \circ m = m' \circ (f \otimes f) \text{ and } f \circ u = u'.$$

A *bialgebra* is an algebra $(\mathcal{B}, m, u)$ and coalgebra $(\mathcal{B}, \Delta, \varepsilon)$ such that Δ and ε are algebra homomorphisms. A bialgebra $\mathcal{B}$ with coproduct Δ is said to be *graded* if it decomposes into submodules $\mathcal{B}_0, \mathcal{B}_1, \ldots$ such that

(1) $\mathcal{B} = \oplus_{n\geq 0}\mathcal{B}_n$,
(2) $\mathcal{B}_i\mathcal{B}_j \subseteq \mathcal{B}_{i+j}$, and
(3) $\Delta(\mathcal{B}_n) \subseteq \oplus_{i+j=n}\mathcal{B}_i \otimes \mathcal{B}_j$.

Definition 2.1. *A bialgebra $(\mathcal{H}, m, u, \Delta, \varepsilon)$ is a* Hopf algebra *if there exists a linear map $S\colon \mathcal{H} \to \mathcal{H}$ (called the* antipode*) such that*

$$m \circ (S \otimes id) \circ \Delta = u \circ \varepsilon = m \circ (id \otimes S) \circ \Delta.$$

A Hopf algebra $\mathcal{H}$ is said to be *connected* if $\mathcal{H}_0 = R$. When the ground ring R is in fact a field K, a *character* (sometimes called a *multiplicative linear functional*) of the Hopf algebra $\mathcal{H}$ is an algebra homomorphism from $\mathcal{H}$ to the field K. A *combinatorial Hopf algebra* is a graded connected Hopf algebra equipped with a character.

Gessel [44] describes an *internal (or inner) coproduct* which takes QSym^n (quasisymmetric functions of degree n) to $\mathrm{QSym}^n \otimes \mathrm{QSym}^n$. This internal coproduct corresponds to the internal coproduct on symmetric functions, taking p_n to $p_n \otimes p_n$. Malvenuto and Reutenauer [90] introduce an *outer coproduct* on QSym defined on the monomial quasisymmetric functions by

$$\Delta(M_{(\beta_1,\beta_2,\ldots,\beta_k)}) = \sum_{i=0}^{k} M_{(\beta_1,\ldots,\beta_i)} \otimes M_{(\beta_{i+1},\ldots,\beta_k)}.$$

For example,

$$\Delta(M_{312}) = 1 \otimes M_{312} + M_3 \otimes M_{12} + M_{31} \otimes M_2 + M_{312} \otimes 1.$$

Restricting this coproduct to symmetric functions takes p_n to $p_n \otimes 1 + 1 \otimes p_n$, and therefore, this coproduct is different from Gessel's internal coproduct. Malvenuto and Reutenauer [90] and Ehrenborg [34] independently discovered the antipode map on QSym (with respect to the outer coproduct), proving that QSym is a Hopf algebra. See [61] for a thorough introduction to the Hopf algebra structure of quasisymmetric functions.

If V is an R-module, let $V^\star := Hom(V, R)$ be its dual R-module. Under certain finiteness conditions (which are true for the situations explored in this article), the duals of Hopf algebras are themselves Hopf algebras. In the language of Hopf algebras and their duality, the Ditters conjecture states that the Leibniz–Hopf algebra is

dual (as a Hopf algebra over the integers) to a free commutative algebra over the integers.

The dual to QSym is the ring (or algebra) of noncommutative symmetric functions, denoted NSym. We take a moment to briefly describe some of the structure of NSym. For a thorough introduction to the topic through the lens of quasideterminants, see [43]; we typically follow their notation conventions.

NSym can be thought of as a free associative algebra $K\langle \Lambda_1, \Lambda_2, \ldots \rangle$ generated by an infinite sequence of noncommuting indeterminates $(\Lambda_k)_{k\geq 1}$ over a fixed field K of characteristic 0. (We usually take K to be $\mathbb{C}$, the complex numbers.) The *noncommutative elementary symmetric functions* are the indeterminates Λ_k, and their generating function is

$$\lambda(t) = \sum_{k\geq 0} t^k \Lambda_k,$$

while the *noncommutative complete homogeneous symmetric functions* S_k are defined by their generating function

$$\sigma(t) = \sum_{k\geq 0} t^k \mathbf{S}_k = \lambda(-t)^{-1}. \tag{2.1}$$

Note that this mirrors the relationship in Sym between the elementary and complete homogeneous symmetric functions, where if $H(t) = \sum_{n\geq 0} h_n t^n$ and $E(t) = \sum_{n\geq 0} e_n t^n$, then $H(t)E(-t) = 1$. Both of these basis analogues are multiplicative, meaning $\mathbf{S}_\alpha = \mathbf{S}_{\alpha_1}\mathbf{S}_{\alpha_2}\cdots\mathbf{S}_{\alpha_k}$ and $\Lambda_\alpha = \Lambda_{\alpha_1}\Lambda_{\alpha_2}\cdots\Lambda_{\alpha_k}$ for $\alpha = (\alpha_1, \alpha_2, \ldots \alpha_k)$. The *Ribbon Schur functions*, which form a basis dual to Gessel's fundamental basis for QSym, can be defined through quasideterminants. Two different candidates for the noncommutative analogue of the power sum symmetric functions will be described in Sect. 4.2. We use boldface letters for bases of NSym, lowercase letters for bases of Sym, and uppercase letters for bases of QSym.

The *forgetful map*, frequently denoted by $\chi\colon$ NSym $\to$ Sym, sends the basis element $\mathbf{S}_\alpha$ to the complete homogeneous symmetric function $h_{\alpha_1}h_{\alpha_2}\cdots h_{\alpha_{\ell(\alpha)}}$. Essentially, the forgetful map "forgets" that the functions don't commute. This map can then be extended linearly to all of NSym and is in fact a surjection (but clearly not a bijection) onto Sym.

Let C be a category of objects. An object T is a *terminal object* for the category C if for all objects $X \in C$ there exists a unique morphism $X \to T$. Not every category has a terminal object, but if such a terminal object exists it is necessarily unique. Aguilar, Bergeron, and Sottile [1] introduce a canonical character ζ_Q on QSym and describe what it does to the monomial and fundamental quasisymmetric functions. Equipped with this character, quasisymmetric functions are the terminal object in the category of combinatorial Hopf algebras.

Theorem 2.2. ([1]) *If $\mathcal{H}$ is a combinatorial Hopf algebra with a character ζ, then there exists a unique homomorphism from $(\mathcal{H}, \zeta)$ to (QSym, ζ_Q) such that the homomorphism on characters induced by the Hopf algebra homomorphism sends ζ to ζ_Q.*

Theorem 2.2 helps to explain why quasisymmetric functions appear in so many different contexts throughout algebraic combinatorics. Examples for which the connection is well-understood include Rota's Hopf algebra of isomorphism classes of finite graded posets [74] and the chromatic Hopf algebra of isomorphism classes of finite unoriented graphs [111]. Note that this mirrors the similar result stating that Sym is the terminal object in the category of cocommutative combinatorial Hopf algebras [1].

2.2 Solomon's Descent Algebra

Let $\mathbb{Z}\mathfrak{S}_n$ be the group ring of permutations $\mathfrak{S}_n$ over the integers and let $\mathbb{Z}\mathfrak{S} = \oplus_{n\geq 0}\mathbb{Z}\mathfrak{S}_n$ be the direct sum of $\mathbb{Z}\mathfrak{S}_n$ over all positive integers n. For each $\sigma \in \mathfrak{S}_n$ let $\mathrm{Des}(\sigma)$ be the *descent set* of σ defined by $\mathrm{Des}(\sigma) = \{i \mid 1 \leq i \leq n-1, \sigma(i) > \sigma(i+1)\}$. To each subset L of $\{1, 2, \ldots, n-1\}$, associate an element D_L of $\mathbb{Z}\mathfrak{S}_n$ as follows:

$$D_L = \sum_{\mathrm{Des}(\sigma)=L} \sigma.$$

The composition $\beta(L) = (L_1, L_2 - L_1, \ldots, n - L_k)$ (where $k = |L|$) is frequently used to index this *descent basis*. For example, let $n = 4$ and $L = \{2\}$. Then

$$D_L = D_{(2,2)} = 1324 + 1423 + 2314 + 2413 + 3412.$$

Let Σ_n be the linear span of the elements D_L and endow $\Sigma := \oplus_{n\geq 0}\Sigma_n \subseteq \mathbb{Z}\mathfrak{S}$ with a ring structure by setting $\sigma\pi = 0$ if $\sigma \in \mathfrak{S}_n$ and $\pi \in \mathfrak{S}_m$ such that $m \neq n$. Σ is called *Solomon's descent algebra*. Solomon [116] proves that Σ is a subalgebra of $\mathbb{Z}\mathfrak{S}$. Gessel [44] shows that the algebra dual to the coalgebra QSym_n (endowed with Gessel's internal coproduct) is isomorphic to Solomon's descent algebra Σ_n.

The set Σ (just as a set, not as the descent algebra) also admits a Hopf algebra structure. That is, Malvenuto and Reutenauer [89, 90] define a product and coproduct on $\mathbb{Z}\mathfrak{S}$ to prove that $\mathbb{Z}\mathfrak{S}$ is a Hopf algebra (called the *Malvenuto–Reutenauer algebra*) with Σ as a Hopf subalgebra. (This algebra is in fact isomorphic to the algebra FQSym of *free quasisymmetric functions*; see [32] for details.) Malvenuto and Reutenauer show (Theorem 3.3 in [90]) that Σ is dual to QSym, with the descent basis $\{D_\alpha\}$ of Σ dual to the basis $\{F_\alpha\}$. This means that the product on one of these bases determines the coproduct on the other, and vice versa. That is, if

$$F_\alpha F_\beta = \sum_{\gamma \models |\alpha|+|\beta|} c^{\gamma}_{\alpha,\beta} F_\gamma,$$

then comultiplication in Σ is defined by

$$\Delta_\Sigma(D_\gamma) = \sum c^{\gamma}_{\alpha,\beta} D_\alpha \otimes D_\beta.$$

This duality pairing also implies that the descent basis is isomorphic to the ribbon Schur basis for noncommutative symmetric functions since the ribbon Schur basis for NSym is dual to the fundamental basis for QSym.

2.3 Representations of the **0**-*Hecke Algebra*

The representation theoretic significance of the fundamental quasisymmetric functions mirrors that of the Schur functions. We first describe the symmetric function case for ease of comparison. Recall that the symmetric group $\mathfrak{S}_n$ can be generated by adjacent transpositions $s_i = (i, i+1)$ for $1 \leq i \leq n-1$ satisfying the following relations:

$$s_i^2 = 1 \text{ for } 1 \leq i \leq n-1,$$

$$s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1} \text{ for } 1 \leq i \leq n-2, \text{ and}$$

$$s_i s_j = s_j s_i \text{ for } |i-j| \geq 2.$$

The Frobenius *characteristic map* is a map from characters of the symmetric group $\mathfrak{S}_n$ to symmetric functions which are homogeneous of degree n. The Schur functions are the images of irreducible characters.

The 0-Hecke algebra is a $\mathbb{C}$-algebra generated by elements satisfying relations similar to the relations on the symmetric group generators described above. That is, $H_n(0)$ is generated by elements $T_1, T_2, \ldots, T_{n-1}$ satisfying:

$$T_i^2 = T_i \text{ for } 1 \leq i \leq n-1,$$

$$T_i T_{i+1} T_i = T_{i+1} T_i T_{i+1} \text{ for } 1 \leq i \leq n-2, \text{ and}$$

$$T_i T_j = T_j T_i \text{ if } |i-j| \geq 2.$$

If σ is a permutation in $\mathfrak{S}_n$ with reduced word $\sigma = s_{i_1} s_{i_2} \cdots s_{i_\ell}$, then define $T_\sigma \in H_n(0)$ by

$$T_\sigma = T_{i_1} T_{i_2} \cdots T_{i_\ell}.$$

The 0-Hecke algebra $H_n(0)$ is a specialization of the Hecke algebra $H_n(q)$ at $q = 0$. (See Méliot [93] for further details on the Hecke algebra $H_n(q)$ and its relationship to the 0-Hecke algebra $H_n(0)$.)

Norton [96] investigates the representation theory of $H_n(0)$ and proves that there are 2^{n-1} distinct irreducible representations of $H_n(0)$, indexed by compositions of n. Let $\mathcal{G}_0(H_n(0))$ be the Grothendieck group of finitely generated $H_n(0)$-modules and $\mathcal{G} = \oplus_{n \geq 0} \mathcal{G}_0(H_n(0))$ be the associated Grothendieck ring. (See Carter [23] for a thorough account of the representation theory of the 0-Hecke algebra and see Huang [71–73] for recent connections with flag varieties, the Stanley–Reisner ring,

and tableaux.) Krob and Thibon [76] prove that $\mathcal{G}$ is isomorphic to the ring of quasisymmetric functions via a characteristic map $\mathcal{F}\colon \mathcal{G} \to \mathrm{QSym}$ called the *quasisymmetric characteristic*. Let L_α denote the irreducible representation of $H_n(0)$ corresponding to α. Then $\mathcal{F}$ sends L_α to the fundamental quasisymmetric function F_α.

Theorem 2.3. ([31]) *The map $\mathcal{F}$, defined by $\mathcal{F}(L_\alpha) = F_\alpha$, is a ring isomorphism between the Grothendieck group $\mathcal{G}$ of finite-dimensional representations of $H_n(0)$ and the ring of quasisymmetric functions.*

The fundamental quasisymmetric functions therefore correspond to characters of irreducible representations of the 0-Hecke algebra. See Hivert [65, 66] to view this through the lens of divided difference operators. Similar representation theoretic interpretations can be ascribed to various other bases for quasisymmetric functions and will be discussed in the relevant sections.

3 Macdonald Polynomials

The Schur functions are uniquely determined by the following two requirements described on p. 305 of Macdonald [88].

(1) Let λ be a partition. Then

$$s_\lambda = m_\lambda + \sum_{\mu < \lambda} K_{\lambda\mu} m_\mu,$$

where $\mu \leq \lambda$ if and only if $\mu_1 + \mu_2 + \cdots + \mu_j \leq \lambda_1 + \lambda_2 + \cdots \lambda_j$ for all j. (This partial ordering is called the *dominance ordering*.) Here the coefficients $K_{\lambda\mu}$ are called the *Kostka numbers*, or *Kostka coefficients*.

(2) $\langle s_\lambda, s_\mu \rangle = \delta_{\lambda\mu}$.

Macdonald [86] generalized this construction to a two-parameter family of functions $P_\lambda = P_\lambda(q,t)$ in $\mathbb{Q}(q,t)$ characterized by the following two requirements.

(1) Let λ be a partition. Then $P_\lambda = m_\lambda +$ lower terms in dominance order.
(2) $\langle P_\lambda, P_\mu \rangle_{q,t} = 0$ if $\lambda \neq \mu$, where

$$\langle p_\lambda, p_\mu \rangle_{q,t} = \delta_{\lambda\mu} z_\lambda \prod_{i=1}^{\ell(\lambda)} \frac{1-q^{\lambda_i}}{1-t^{\lambda_i}}.$$

Notice that when $q = t$, the scalar product reduces to $\langle p_\lambda, p_\mu \rangle_{q,q} = \delta_{\lambda\mu} z_\lambda$ and so $P_\lambda(q,q) = s_\lambda$. Similarly, $P_\lambda(q,1) = m_\lambda$ and $P_\lambda(1,t) = e_{\lambda'}$. (See p. 324 of Macdonald [88].) Therefore, the Macdonald polynomials are simultaneous generalizations of several different symmetric function bases. Macdonald polynomials also appear in connection with the Hilbert scheme of n points in the plane [59].

There are several variations on the original definition of Macdonald polynomials, including the *modified Macdonald polynomials*, $\tilde{H}_\mu$, obtained from P_μ by certain substitutions and motivated by their connection to the coefficients appearing in the Schur function expansion of Macdonald polynomials. Haglund conjectured and Haglund, Haiman, and Loehr proved [50, 51, 54, 55] a combinatorial formula for the Macdonald polynomials $\tilde{H}_\mu$ using statistics on fillings of partition diagrams. To describe this formula, we introduce several pertinent definitions.

Recall that a filling $\sigma\colon \mu \to \mathbb{Z}^+$ is a function from the cells of the diagram of a partition μ to the positive integers. The *reading word* of the filling is the word obtained by reading the entries of the filling from top to bottom, left to right.

The major index and inversion statistic on permutations can be generalized to statistics on fillings of Ferrers diagrams. Let s be a cell in the partition diagram μ and let South(s) be the cell immediately below s in the same column as s. Define

$$\mathrm{Des}(\sigma, \mu) = \{s \in \mu \mid \sigma(s) > \sigma(\mathrm{South}(s))\}.$$

(No cell in the bottom row of μ can be in Des(σ, μ).) Let leg(s) be the number of cells above s in the same column as s and let $arm(s)$ be the number of cells to the right of s in the same row as s. Then

$$\mathrm{maj}(\sigma, \mu) = \sum_{s \in \mathrm{Des}(\sigma,\mu)} (\mathrm{leg}(s) + 1).$$

Let u, v, w be three cells in the diagram of μ such that u and v are in the same row of μ with v strictly to the right of u and $w = \mathrm{South}(u)$ as shown:

$$\begin{array}{|c|c|c|}\hline u & \cdots & v \\ \hline w & \multicolumn{2}{c}{} \\ \cline{1-1}\end{array}.$$

Any collection of three cells arranged in this way is called a *triple*. Define an orientation on the cells in a triple of a filling σ of μ by starting with the cell containing the smallest entry and moving in a circular motion from smallest to largest. (If two entries are equal, the one which appears first in the reading word is considered smaller.) If the resulting orientation is counterclockwise, the triple is called an *inversion triple* (see Fig. 7). Two cells u, v in the bottom row are also considered an inversion triple if v is strictly to the right of u and $\sigma(u) > \sigma(v)$. The total number of inversion triples in a filling σ of a partition μ is denoted inv(σ, μ).

For example, the filling in Fig. 8 has descent set Des $= \{(1, 3), (2, 3), (2, 2)\}$, where cells are indexed by (column, row) to mimic the (x, y) Cartesian coordinates.

$$\begin{array}{|c|c|c|}\hline 3 & \cdots & 2 \\ \hline 1 & \multicolumn{2}{c}{} \\ \cline{1-1}\end{array} \qquad \begin{array}{|c|c|c|}\hline 2 & \cdots & 1 \\ \hline 2 & \multicolumn{2}{c}{} \\ \cline{1-1}\end{array} \qquad \begin{array}{|c|c|c|}\hline 1 & \cdots & 2 \\ \hline 2 & \multicolumn{2}{c}{} \\ \cline{1-1}\end{array} \qquad \begin{array}{|c|c|c|}\hline 4 & \cdots & 1 \\ \hline 3 & \multicolumn{2}{c}{} \\ \cline{1-1}\end{array}$$

Fig. 7 First two triples are inversion triples; the third and fourth are not

Fig. 8 A filling of the partition (4, 3, 2) with reading word 582613371

5	8		
2	6	1	
3	3	7	1

The major index for this filling is $1 + 1 + 2 = 4$ and $\mathrm{inv}(\sigma, \mu) = 3 + 2 = 5$, since there are three inversion triples in the bottom row and two additional inversion triples.

Theorem 3.1. ([54, 55]) *Let μ be a partition of n. Then*

$$\tilde{H}_\mu(X; q, t) = \sum_{\sigma:\, \mu \to \mathbb{Z}^+} x^\sigma q^{inv(\sigma,\mu)} t^{maj(\sigma,\mu)}.$$

Theorem 3.1 provides a straightforward method for computing Macdonald polynomials. This formula could potentially be used to find a product rule for Macdonald polynomials utilizing tableau constructions, although Yip recently found an elegant combinatorial rule for multiplying Macdonald polynomials [124] using the alcove walk model introduced by Ram and Yip [104].

3.1 Quasisymmetric Decomposition of Macdonald Polynomials

Macdonald polynomials can also be described as sums of fundamental quasisymmetric functions with coefficients in q and t.

Theorem 3.2. ([52, 54]) *Let μ be a partition of n. Then*

$$\tilde{H}_\mu(X; q, t) = \sum_{\beta \in \mathfrak{S}_n} q^{inv(\beta,\mu)} t^{maj(\beta,\mu)} F_{Des(\beta^{-1})},$$

where each permutation β in the sum corresponds to the standard filling of μ with reading word β and $\mathrm{Des}(\beta^{-1})$ *is the usual descent set of the permutation β^{-1} obtained by taking the inverse of β.*

For example, if $\mu = (2, 1)$, the following Table demonstrates that

$$\tilde{H}_{21}(X; q, t) = F_3 + (q + t)F_{21} + (q + t)F_{12} + qtF_{111}.$$

This expansion of the Macdonald polynomials into fundamental quasisymmetric functions paves the way for new approaches to long-standing open questions. For example, Macdonald [86] conjectured that the coefficients in the expansion of $\tilde{H}_\mu$ into Schur functions are polynomials in q and t with nonnegative integer coefficients. Haiman [59] proved this by showing that $\tilde{H}_\mu$ is the bigraded Frobenius character of

Permutation β (Reading word of filling)	123	132	213	231	312	321
Filling of μ	$\begin{array}{cc} 1 & \\ 2 & 3 \end{array}$	$\begin{array}{cc} 1 & \\ 3 & 2 \end{array}$	$\begin{array}{cc} 2 & \\ 1 & 3 \end{array}$	$\begin{array}{cc} 2 & \\ 3 & 1 \end{array}$	$\begin{array}{cc} 3 & \\ 1 & 2 \end{array}$	$\begin{array}{cc} 3 & \\ 2 & 1 \end{array}$
$\mathrm{inv}(\beta, \mu)$	0	1	0	1	0	1
$\mathrm{maj}(\beta, \mu)$	0	0	1	0	1	1
β^{-1}	123	132	213	312	231	321
$\mathrm{Des}(\beta^{-1})$	$\emptyset$	2	1	1	2	1, 2

a doubly graded $\mathfrak{S}_n$-module, but this approach did not provide an explicit combinatorial formula for the coefficients. Assaf's dual equivalence [6, 7] provides another potential approach to Schur positivity which makes use of the decomposition of a symmetric function into fundamental quasisymmetric functions.

The Hall–Littlewood polynomials are a one-parameter specialization of Macdonald polynomials introduced by Littlewood as a symmetric function realization of the Hall algebra [82]. Several different candidates for quasisymmetric Hall–Littlewood polynomials have recently been proposed. See Hivert [66] for an analogue in NSym and its QSym companion, and see Novelli, Thibon, and Williams [97] for a different noncommutative analogue. Connections between these two approaches are studied in Novelli, Tevlin, and Thibon [98]. See also Haglund, Luoto, Mason, and van Willigenburg [57] for another quasisymmetric analogue.

3.2 Quasisymmetric Schur Functions

Haglund's formula (Theorem 3.1) to generate the Macdonald polynomials using statistics on fillings of partition diagrams is generalized in [56] to fillings of weak composition diagrams in order to generate the nonsymmetric Macdonald polynomials introduced and initially developed by Cherednik [24], Macdonald [87], Opdam [99], and Sahi [109]. When these polynomials are specialized to $q = t = \infty$, the resulting polynomials, called *Demazure atoms* due to their connections to Demazure characters, form a basis for all polynomials. The Demazure atoms decompose the Schur functions in a natural way, and their generating diagrams satisfy a Robinson–Schensted–Knuth-style algorithm [91]. Type A key polynomials [78, 105] are positive sums of Demazure atoms [92]. Summing the Demazure atoms over all weak compositions which collapse to a fixed composition when their zeros are removed produces a new collection of quasisymmetric functions, called the *quasisymmetric Schur functions*, which we now formally define using fillings of composition diagrams [57].

Let α be a composition of n. If T is a filling of the composition diagram α (written in English notation) satisfying the following properties, then T is called a *semi-standard reverse composition tableau*, abbreviated SSRCT.

(1) The entries in each row weakly decrease when read from left to right.

(2) The entries in the leftmost column strictly increase when read from top to bottom.
(3) (Triple Rule) If $k > j$ and $T(i,k) \geq T(i+1,j)$ (for cells (i,k) and $(i+1,j)$), then $(i+1,k)$ is a cell in α and $T(i+1,k) > T(i+1,j)$. (Here, if there is no cell at coordinate (i,j), set $T(i,j) = 0$.)

The set of all semi-standard reverse composition tableaux of shape α is denoted $\mathrm{SSRCT}(\alpha)$. The *weight* of a semi-standard reverse composition tableau T, denoted X^T, is the product over all i of $x_i^{\#(i)}$, where $\#(i)$ is the number of times i appears in T.

Definition 3.3. *The* quasisymmetric Schur function $\mathcal{S}_\alpha$ *is defined by*

$$\mathcal{S}_\alpha(X) = \sum_{T \in SSRCT(\alpha)} X^T.$$

Quasisymmetric Schur functions form a basis for QSym and are closely related to Schur functions. In fact, the quasisymmetric Schur functions, when summed over all rearrangements of a given partition, produce the Schur function indexed by this partition [57]. That is,

$$s_\lambda = \sum_{\tilde{\alpha}=\lambda} \mathcal{S}_\alpha,$$

where $\tilde{\alpha}$ is the partition obtained by arranging the parts of α into weakly decreasing order.

For example, the four semi-standard reverse composition tableaux of shape $(2, 1)$ are

$$\begin{ytableau} 1 & 1 \\ 2 \end{ytableau}, \quad \begin{ytableau} 1 & 1 \\ 3 \end{ytableau}, \quad \begin{ytableau} 2 & 1 \\ 3 \end{ytableau}, \quad \text{and} \quad \begin{ytableau} 2 & 2 \\ 3 \end{ytableau},$$

producing the quasisymmetric Schur function

$$\mathcal{S}_{21}(x_1, x_2, x_3) = x_1^2x_2 + x_1^2x_3 + x_1x_2x_3 + x_2^2x_3.$$

The four semi-standard reverse composition tableaux of shape $(1, 2)$ are

$$\begin{ytableau} 1 \\ 2 & 2 \end{ytableau}, \quad \begin{ytableau} 1 \\ 3 & 2 \end{ytableau}, \quad \begin{ytableau} 1 \\ 3 & 3 \end{ytableau}, \quad \text{and} \quad \begin{ytableau} 2 \\ 3 & 3 \end{ytableau},$$

producing the quasisymmetric Schur function

$$\mathcal{S}_{12}(x_1, x_2, x_3) = x_1x_2^2 + x_1x_2x_3 + x_1x_3^2 + x_2x_3^2.$$

Together these sum to $s_{21}(x_1, x_2, x_3)$; that is,

$$s_{21} = \mathcal{S}_{21} + \mathcal{S}_{12}.$$

The quasisymmetric Schur functions expand positively in the fundamental basis for QSym. The set of all *standard reverse composition tableaux* of shape α, abbreviated SRCT(α), is the subset of SSRCT(α) consisting of the semi-standard reverse composition tableaux in which each of the positive integers in the set $\{1, 2, \ldots, |\alpha|\}$ appears exactly once. Each standard reverse composition tableau T has a *descent set* $\mathrm{Des}(T)$ defined by

$$\mathrm{Des}(T) = \{i \mid i+1 \text{ appears weakly right of } i\} \subseteq [n-1].$$

Theorem 3.4. ([57]) *The quasisymmetric Schur function $\mathcal{S}_\alpha$ decomposes into the fundamental basis for quasisymmetric functions as follows:*

$$\mathcal{S}_\alpha = \sum_{T \in SRCT(\alpha)} F_{Des(T),|\alpha|}.$$

For example, the three standard reverse composition tableaux of shape $(2, 1, 3)$ are

$$\begin{array}{|c|c|cc} \hline 3 & 1 & \multicolumn{1}{|c}{} & \\ \hline 4 & \multicolumn{1}{|c}{} & & \\ \hline 6 & 5 & \multicolumn{1}{c|}{2} & \\ \hline \end{array}\;, \quad \begin{array}{|c|c|cc} \hline 2 & 1 & \multicolumn{1}{|c}{} & \\ \hline 4 & \multicolumn{1}{|c}{} & & \\ \hline 6 & 5 & \multicolumn{1}{c|}{3} & \\ \hline \end{array}\;, \quad \text{and} \quad \begin{array}{|c|c|cc} \hline 2 & 1 & \multicolumn{1}{|c}{} & \\ \hline 3 & \multicolumn{1}{|c}{} & & \\ \hline 6 & 5 & \multicolumn{1}{c|}{4} & \\ \hline \end{array}\;.$$

The descent sets are, respectively, $\{1, 3, 4\}$, $\{2, 4\}$, and $\{2, 3\}$. This implies that

$$\mathcal{S}_{213} = F_{\{1,3,4\},6} + F_{\{2,4\},6} + F_{\{2,3\},6}.$$

Tewari and van Willigenburg [123] introduce a collection of operators $\{\pi_i\}_{i=1}^{n-1}$ on standard reverse composition tableaux (which satisfy the same relations as the generators $\{T_i\}_{i=1}^{n-1}$ described in Sect. 2.3) to produce an $H_n(0)$-action on standard reverse composition tableaux of size n.

In particular, for $T \in \mathrm{SRCT}(\alpha)$ for some composition $\alpha \models n$ and $1 \le i \le n-1$, entries i and $i+1$ are said to be *attacking* if they are in the same column of T or they are in adjacent columns of T with $i+1$ appearing to the right of i in a strictly lower row. The operators π_i for $1 \le i \le n-1$ are defined as follows, where $s_i(T)$ interchanges the positions of entries i and $i+1$.

$$\pi_i(T) = \begin{cases} T & \text{if } i \notin \mathrm{Des}(T), \\ 0 & \text{if } i \in \mathrm{Des}(T),\ i \text{ and } i+1 \text{ are attacking, and} \\ s_i(T) & \text{if } i \in \mathrm{Des}(T),\ i \text{ and } i+1 \text{ are nonattacking.} \end{cases}$$

Extend these operators to all of $\mathfrak{S}_n$ by setting $\pi_\sigma = \pi_{i_1}\pi_{i_2}\cdots\pi_{i_\ell}$ when $\sigma = s_{i_1}s_{i_2}\cdots s_{i_\ell}$ is any reduced word for σ. Define a partial order $\preceq_\alpha$ on SRCT(α) by setting $T_1 \preceq_\alpha T_2$ if and only if $\pi_\sigma(T_1) = T_2$ for some permutation $\sigma \in \mathfrak{S}_n$. Extend $\preceq_\alpha$ to a total order $\preceq_\alpha^t$ arbitrarily and let $\mathcal{V}_{T_i}$ be the $\mathbb{C}$-linear span of all $T_j \in \mathrm{SRCT}(\alpha)$ such that $T_j \succeq_\alpha^t T_i$.

Theorem 3.5. ([123]) *If $T_1 \in SRCT(\alpha)$ is the minimal element under the total order $\preceq^t_\alpha$, then $\mathcal{V}_{T_1} := \mathbf{S}$ is an $H_n(0)$-module whose quasisymmetric characteristic is the quasisymmetric Schur function $\mathcal{S}_\alpha$.*

A *simple* composition is a composition $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_\ell)$ such that if $\alpha_i \geq \alpha_j \geq 2$ and $1 \leq i < j \leq \ell$, then there exists an integer k satisfying $i < k < j$ such that $\alpha_k = \alpha_j - 1$.

Tewari and van Willigenburg [123] prove that $\mathbf{S}$ is an indecomposable $H_n(0)$-module if and only if α is simple. These results lead to the introduction of a new basis for quasisymmetric functions called the *canonical quasisymmetric functions* $\{\mathcal{C}_\alpha\}_\alpha$ and a branching rule for the $\mathbf{S}$ which is analogous to the classical branching rule for Schur functions [108].

The product of a quasisymmetric Schur function and a Schur function expands into the quasisymmetric Schur function basis through a rule which refines the Littlewood–Richardson Rule [58] but a formula for the coefficients appearing in the product of arbitrary quasisymmetric Schur functions is unknown. See [85] for a thorough introduction to quasisymmetric Schur functions and their closely related counterpart, the *Young quasisymmetric Schur functions*. The Young quasisymmetric Schur functions are obtained from the quasisymmetric Schur functions by a simple reversal of the indexing composition and the variables, but at times the Young quasisymmetric Schur functions are easier to work with due to their compatibility with semi-standard Young tableaux (rather than reverse semi-standard Young tableaux).

4 Quasisymmetric Analogues of Symmetric Function Bases

Quasisymmetric functions play a major role in answering important questions about symmetric functions. Analogues in QSym of classical bases for symmetric functions aid in this pursuit by providing a dictionary to translate between Sym and QSym. We have already discussed a quasisymmetric analogue of the monomial symmetric functions as well as two different quasisymmetric analogues of the Schur functions. We now introduce another natural quasisymmetric analogue of the Schur function basis as well as a quasisymmetric analogue of the power sum basis.

4.1 Dual Immaculate Quasisymmetric Functions

Berg, Bergeron, Saliola, Serrano, and Zabrocki [12] generalize Bernstein's creation operator construction of the Schur functions to obtain a basis for NSym called the *immaculate basis* and denoted $\mathcal{I}_\alpha$. The dual basis in QSym, called the *dual immaculate quasisymmetric functions*, can be generated by fillings of tableaux as follows.

Let $F : \alpha \to \mathbb{Z}^+$ be a filling of a composition diagram α with positive integers such that the sequence of entries in each row (read from left to right) is weakly

$$\begin{array}{|c|c|c|}\hline 4 & 4 & 5 \\ \hline 3 & \multicolumn{2}{c}{} \\ \cline{1-2} 1 & 2 & \multicolumn{1}{c}{} \\ \cline{1-2}\end{array} \qquad \begin{array}{|c|c|c|}\hline 4 & 4 & 5 \\ \hline 2 & \multicolumn{2}{c}{} \\ \cline{1-2} 1 & 3 & \multicolumn{1}{c}{} \\ \cline{1-2}\end{array} \qquad \begin{array}{|c|c|c|}\hline 3 & 4 & 4 \\ \hline 2 & \multicolumn{2}{c}{} \\ \cline{1-2} 1 & 5 & \multicolumn{1}{c}{} \\ \cline{1-2}\end{array} \qquad \begin{array}{|c|c|c|}\hline 3 & 4 & 5 \\ \hline 2 & \multicolumn{2}{c}{} \\ \cline{1-2} 1 & 4 & \multicolumn{1}{c}{} \\ \cline{1-2}\end{array}$$

Fig. 9 Four immaculate tableaux of shape (2, 1, 3) and weight $x_1x_2x_3x_4^2x_5$

increasing and the sequence of entries in the leftmost column (read from bottom to top) is strictly increasing. Then F is said to be an *immaculate tableau* of shape α. (Note that since we are using French notation our definition varies slightly from the definition in [12] but produces the same diagrams modulo a horizontal flip.) The *weight* of an immaculate tableau U, denoted x^U, is the product over all i of $x_i^{\#(i)}$, where $\#(i)$ is the number of times i appears in U.

Definition 4.1. ([12]) *Let α be a composition. The* dual immaculate quasisymmetric function $\mathcal{I}_\alpha^*$ *is given by*

$$\mathcal{I}_\alpha^* = \sum_U x^U,$$

where the sum is over all immaculate tableaux of shape α.

For example, the coefficient of $x_1x_2x_3x_4^2x_5$ in $\mathcal{I}_{2,1,3}^*$ is 4 since there are four immaculate tableaux of shape (2, 1, 3) and weight $x_1x_2x_3x_4^2x_5$. (These immaculate tableaux are given in Fig. 9.)

The following theorem provides a formula for the expansion of the Schur functions into the dual immaculate quasisymmetric functions.

Theorem 4.2. ([12]) *Let λ be a partition of length k. Then*

$$s_\lambda = \sum_\sigma (-1)^\sigma \mathcal{I}^*_{\lambda_{\sigma_1}+1-\sigma_1, \lambda_{\sigma_2}+2-\sigma_2, \ldots, \lambda_{\sigma_k}+k-\sigma_k},$$

where $(-1)^\sigma$ is the sign of σ and the sum is over all permutations $\sigma \in \mathfrak{S}_k$ such that $\lambda_{\sigma_i} + i - \sigma_i > 0$ for all $1 \le i \le k$.

For example,

$$s_{321} = \mathcal{I}_{321}^* - \mathcal{I}_{141}^*.$$

Note that the coefficients are not always nonnegative and further the compositions indexing the terms appearing in this expansion are not merely rearrangements of the partition λ as is the case in the quasisymmetric Schur expansion of the Schurs. However, the beauty of the connection to Schur functions is more readily apparent in the dual, since applying the forgetful map to an immaculate function produces the corresponding Schur function. That is, $\chi(\mathcal{I}_\alpha) = s_\alpha$.

Grinberg recently proved Zabrocki's conjecture that the dual immaculate quasisymmetric functions can also be constructed using a variation on Bernstein's creation operators [46]. The dual immaculate quasisymmetric functions expand into

positive sums of the monomial quasisymmetric functions, the fundamental quasisymmetric functions, and, recently shown in [3], the Young quasisymmetric Schur functions. The latter expansion is not at all obvious given the very different methods used to generate these two bases and therefore provides further justification that both of these families of functions are interesting and natural objects of study.

Like quasisymmetric Schur functions, dual immaculate quasisymmetric functions correspond to characteristics of certain representations of the 0-Hecke algebra [13], but for the dual immaculate quasisymmetric functions these representations are indecomposable. In particular, let $\mathcal{M}_\alpha$ be the vector space spanned by all words on the letters $\{1, 2, \ldots, \ell(\alpha)\}$ such that the letter j appears α_j times. Define an action of the 0-Hecke algebra on words by

$$\pi_i(w) = \begin{cases} w & w_i \geq w_{i+1} \\ s_i(w) & w_i < w_{i+1}, \end{cases}$$

where $s_i(w) = w_1 w_2 \cdots w_{i-1} w_{i+1} w_i w_{i+2} \cdots w_n$. Note that this is isomorphic to the induced representation

$$\mathrm{Ind}_{H_{\alpha_1}(0)\otimes H_{\alpha_2}(0)\otimes\cdots\otimes H_{\alpha_{\ell(\alpha)}}(0)}^{H_n(0)} (L_{\alpha_1} \otimes L_{\alpha_2} \otimes \cdots \otimes L_{\alpha_m}),$$

where L_k is the one-dimensional representation indexed by the composition (k). A word w in which the first instance of j appears before the first instance of $j+1$ is called a $\mathcal{Y}$-word. The 0-Hecke action defined above cannot move a $j+1$ to the right of a j, so the subspace $\mathcal{N}_\alpha$ of $\mathcal{M}_\alpha$ spanned by all words which are not $\mathcal{Y}$-words is a submodule of $\mathcal{M}_\alpha$.

Theorem 4.3. ([13]) *The characteristic of* $\mathcal{V}_\alpha := \mathcal{M}_\alpha/\mathcal{N}_\alpha$ *is the dual immaculate quasisymmetric function indexed by* α. *In other words,* $\mathcal{F}([\mathcal{V}_\alpha]) = \mathcal{I}^*_\alpha$.

Bergeron, Sánchez-Ortega, and Zabrocki found a Pieri rule (first conjectured in [12] and proved in [18]) for the product of a fundamental quasisymmetric function and a dual immaculate quasisymmetric function, and much is known about the multiplication of the immaculate basis. However, multiplication rules in full generality for the dual immaculate quasisymmetric functions are still largely unknown.

4.2 Quasisymmetric Analogues of the Power Sum Basis

The power sum symmetric functions (defined in Sect. 1.1) are eigenvectors for the omega involution ω; that is, $\omega(p_\lambda) = \varepsilon_\lambda p_\lambda$, where $\varepsilon_\lambda = (-1)^{n-\ell(\lambda)}$ [119]. Power sum symmetric functions are also helpful in computing characters of the symmetric group via the Murnaghan–Nakayama rule [94, 95].

Malvenuto and Reutenauer [90], through the Hopf algebraic dual, NSym, of QSym, introduce a quasisymmetric analogue of the power sum symmetric functions,

also obtained independently by Derksen [29] using a similar process but with a computational error which leads to a different formula. To understand their construction, we recall several facts about generating functions for symmetric and noncommutative symmetric functions. The complete homogeneous symmetric functions, elementary symmetric functions, and power sum symmetric functions (in n variables) can be defined through their generating functions

$$H(t) = \sum_{d \geq 0} h_d t^d = \prod_{i=1}^{n} \frac{1}{1 - x_i t}\ , \quad E(t) = \sum_{k \geq 0} e_k t^k = \prod_{i=1}^{n} (1 + x_i t), \text{ and } P(t) = \sum_{k \geq 1} p_k \frac{t^k}{k}.$$

The relationship between these is given by Newton's formula:

$$-\frac{d}{dt}(E(-t)) = P(t)E(t),$$

which is equivalent to

$$\frac{d}{dt}(H(t)) = H(t)P(t).$$

In their seminal work on noncommutative symmetric functions, Gelfand, Krob, Lascoux, Leclerc, Retakh, and Thibon [43] define a noncommutative analogue of the complete homogeneous symmetric functions (denoted $\mathbf{S}_k$) by describing their generating function (see Sect. 2.1) and requiring they satisfy the multiplicative property.

They then utilize this approach to construct two analogues of the power sums in NSym by requiring that the generating functions satisfy the appropriate analogues of Newton's formula. *Noncommutative power sum symmetric functions of the first kind*, denoted Ψ_k, are defined by

$$\psi(t) = \sum_{k \geq 1} t^{k-1} \Psi_k, \quad \frac{d}{dt}\sigma(t) = \sigma(t)\psi(t), \quad \text{and} \quad \Psi_\alpha = \Psi_{\alpha_1} \Psi_{\alpha_2} \cdots \Psi_{\alpha_\ell},$$

where $\sigma(t)$ is as defined in Eq. 2.1. Similarly, *noncommutative power sum symmetric functions of the second kind*, denoted Φ_k, are defined by

$$\sigma(t) = \exp\left(\sum_{k \geq 1} t^k \frac{\Phi_k}{k}\right) \text{ and } \Phi_\alpha = \Phi_{\alpha_1} \Phi_{\alpha_2} \cdots \Phi_{\alpha_\ell}.$$

Taking the Hopf algebraic duals of these noncommutative power sum bases produces two different quasisymmetric analogues of power sums. We use Ψ and Φ as notation for these to emphasize their relationship with their noncommutative duals. The dual of the noncommutative power sum basis of the first kind is defined [11] by

$$\Psi_\alpha = z_\alpha \sum_{\beta \preceq \alpha} \frac{M_\beta}{\pi(\alpha, \beta)},$$

where the ordering used is the refinement partial order (so that $\alpha \succeq \beta$ if α is coarser than β) and $\pi(\alpha, \beta)$ is given by the following process. First define $\pi(\alpha) = \prod_{i=1}^{\ell(\alpha)} \sum_{j=1}^{i} \alpha_j$. Then for α a refinement of β, set $\pi(\alpha, \beta) = \prod_{i=1}^{\ell(\beta)} \pi(\alpha^{(i)})$, where $\alpha^{(i)}$ consists of the parts of α that combine to β_i.

For example, $\Psi_{312} = (1 \cdot 2 \cdot 3)(\frac{1}{3 \cdot 1 \cdot 2} M_{312} + \frac{1}{3 \cdot 4 \cdot 2} M_{42} + \frac{1}{3 \cdot 1 \cdot 3} M_{33} + \frac{1}{3 \cdot 4 \cdot 6} M_6)$, which simplifies to

$$\Psi_{312} = M_{312} + \frac{1}{4} M_{42} + \frac{2}{3} M_{33} + \frac{1}{12} M_6.$$

Similarly, a formula for quasisymmetric power sums of the second kind is also given in terms of the monomial quasisymmetric functions.

$$\Phi_\alpha = \sum_{\alpha \succeq \beta} \frac{M_\beta}{f(\alpha, \beta)},$$

where the ordering used is again the refinement partial order, and the function $f(\alpha, \beta)$ is given by the following process. Assume $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$. Write α as a concatenation $\alpha^{(1)}\alpha^{(2)} \cdots \alpha^{(k)}$ of compositions $\alpha^{(i)}$ where $\alpha^{(i)} \models \beta_i$. Then $f(\alpha, \beta) = \ell(\alpha^{(1)})! \cdots \ell(\alpha^{(k)})!$.

For example, $\Phi_{312} = (1 \cdot 2 \cdot 3)(\frac{1}{1 \cdot 1 \cdot 1} M_{312} + \frac{1}{2 \cdot 1} M_{42} + \frac{1}{1 \cdot 2} M_{33} + \frac{1}{6} M_6)$, which simplifies to

$$\Phi_{312} = 6M_{312} + 3M_{42} + 3M_{33} + M_6.$$

This formula differs from that of Malvenuto and Reutenauer [90] (who use the notation P_α instead of Φ_α) only by a constant. This constant ensures that the Φ_α refine the symmetric power sums so that

$$p_\lambda = \sum_{\tilde{\alpha} = \lambda} \Phi_\alpha,$$

which is not true for the P_α.

The reader might wonder about the duals of the elementary and complete homogeneous symmetric functions. In fact, the noncommutative complete homogeneous symmetric functions are dual to the monomial quasisymmetric functions, while the noncommutative elementary symmetric functions are dual to the "forgotten" basis for quasisymmetric functions, whose combinatorial structure is largely unknown.

Recall that the fundamental quasisymmetric functions satisfy the following relationship to monomial quasisymmetric functions:

$$F_\alpha = \sum_{\beta \preceq \alpha} M_\beta,$$

where $\beta \preceq \alpha$ again means that β is a refinement of α. Hoffman [67] studied a variation on the fundamental basis called the *essential quasisymmetric functions* E_α, obtained by reversing the inequality in the above equation so that

$$E_\alpha = \sum_{\beta \succeq \alpha} M_\beta.$$

Summing over all coarsenings of α is a natural thing to do because of what the antipode map does to the monomial quasisymmetric functions:

$$S(M_\alpha) = (-1)^{\ell(\alpha)} E_\alpha.$$

Multiplication in the essential basis follows the same rules (modulo a sign) as multiplication in the monomial basis.

4.3 The Shuffle Algebra

The shuffle algebra is a Hopf algebra (whose multiplicative structure is given by an operation called a *shuffle*) which is in fact isomorphic as a graded Hopf algebra to QSym (over the rationals). More details on the shuffle algebra and the closely related concept of Lyndon words can be found in Reutenauer [107], Lothaire [84], or Grinberg-Reiner [47].

Let A be a totally ordered set, which we will call an *alphabet*. A *word* of length n is an ordered string $w = w_1 w_2 \cdots w_n$ of elements of A. Let A^* be the set of all words on the alphabet A. For the purposes of this section, we will take the alphabet to be the positive integers, as is done in [60]. When A is taken to be the positive integers, the *degree* ($|u|$) of a word u in A^* is the sum of its letters rather than the number of letters. The shuffle, $u \sqcup\!\sqcup v$, of two words $u = u_1 u_2 \cdots u_k$ and $v = v_1 v_2 \cdots v_\ell$ in A^* is the sum of all words in A^* of length $k + \ell$ formed from the letters of u and the letters of v such that for all i, u_i appears before u_{i+1} and v_i appears before v_{i+1}. Multiplicities will occur if a letter appears in both u and v. If a letter appears more than once within one of the words u or v, simply consider each occurrence as a distinct letter by applying a different subscript to each appearance of a given letter. This product is associative and can therefore be extended to the shuffle product of a finite number of words. (It is called a shuffle because it resembles the interleaving method used to shuffle a deck of cards.) For example,

$$23 \sqcup\!\sqcup 12 = 2312 + 2132 + 2123 + 1223 + 1223 + 1232.$$

Shuffles in fact guide the multiplication of quasisymmetric power sums of both types. Let a_j equal the number of parts of size j in α, b_j equal the number of parts of

size j in β, and let $\alpha \cdot \beta$ denote their concatenation. Define $C(\alpha, \beta) = \prod_j \binom{a_j+b_j}{a_j}$, so that $C(\alpha, \beta) = z_{\alpha \cdot \beta}/(z_\alpha z_\beta)$.

Theorem 4.4. ([11]) *Let α and β be compositions. Then*

$$\Psi_\alpha \Psi_\beta = \frac{1}{C(\alpha, \beta)} \sum_{\gamma \in \alpha \sqcup\!\sqcup\, \beta} \Psi_\gamma \qquad \text{and} \qquad \Phi_\alpha \Phi_\beta = \frac{1}{C(\alpha, \beta)} \sum_{\gamma \in \alpha \sqcup\!\sqcup\, \beta} \Phi_\gamma.$$

The *shuffle algebra* $K\langle A \rangle$ (where K is a commutative ring with unit) is the set of all linear combinations over K of words on an alphabet A, endowed with this shuffle product. There are a number of distinct proofs that this algebra is isomorphic to QSym, including those of Hazewinkel [60] and Hazewinkel-Gubareni-Kirichenko [63]. Note that Theorem 4.4 in fact implies that the shuffle algebra is isomorphic to QSym. To see this, distribute the $C(\alpha, \beta)$ in the first equation in Theorem 4.4 so that:

$$\frac{\Psi_\alpha}{z_\alpha} \frac{\Psi_\beta}{z_\beta} = \sum_{\gamma \in \alpha \sqcup\!\sqcup\, \beta} \frac{\Psi_\gamma}{z_{\alpha \cdot \beta}}.$$

Then map from the shuffle algebra to QSym via the map $\alpha \mapsto \frac{\Psi_\alpha}{z_\alpha}$. Extend this map linearly to an isomorphism between the shuffle algebra and QSym.

We now discuss the algebraic structure of QSym. First we shall see that QSym over the rationals is a polynomial algebra in the quasisymmetric power sums. Then we describe Hazewinkel's polynomial generators for QSym over the integers.

Let $w \in A^*$ be a word on the alphabet A. Then a *proper suffix* of w is a word $v \in A^*$ such that there exists a nonempty $u \in A^*$ such that $w = uv$. A *prefix* of w is a word $u \in A^*$ such that $w = uv$. Let $\leq_A$ be a total ordering on A^* defined as follows. Let $u = u_1 u_2 \cdots u_k$ and $v = v_1 v_2 \cdots v_m$. If u is a prefix of v then $u \leq_A v$. Otherwise let j be the smallest positive integer such that $u_j \neq v_j$. If $u_j > v_j$ then $u >_A v$. Otherwise $u <_A v$.

Definition 4.5. *A* Lyndon word *is a nonempty word $w \in A^*$ such that every nonempty proper suffix v of w satisfies $w <_A v$. Let $\mathcal{L}$ denote the set of all Lyndon words.*

For example, the words 1324, 1323, and 11213 are Lyndon words while the words 4132, 3241, 2332, and 2233 are not. The shuffle algebra is freely generated over the rationals by the Lyndon words.

Theorem 4.6. ([103]) *Every element of $K\langle A \rangle$ can be uniquely expressed as a polynomial in the Lyndon words. In other words, the shuffle algebra $K\langle A \rangle$ is the polynomial algebra in the Lyndon words.*

One can think of Theorem 4.6 (commonly known as *Radford's Theorem*) as the statement that for any vector space basis whose elements are indexed by words in A^* and whose multiplication is given by shuffles, each basis element can be written as a polynomial in basis elements indexed by Lyndon words.

For example, $w = 321$ is not a Lyndon word, but Ψ_{321} can be expressed as a polynomial in quasisymmetric power sums indexed by Lyndon words. That is,

$$\Psi_{321} = \Psi_1 \cdot \Psi_2 \cdot \Psi_3 - \Psi_{23} \cdot \Psi_1 - \Psi_3 \cdot \Psi_{12} + \Psi_{123}.$$

Radford's theorem implies that quasisymmetric power sums indexed by Lyndon words form an algebraically independent generating set for QSym over the rationals (see [11, 47, 90, 107] for further details).

We now describe Hazewinkel's polynomial generators for QSym, which are indexed by a subset of Lyndon words. The ring $\mathbb{Z}[x_1, x_2, \dots,]$ is endowed with a well-known λ-ring structure via

$$\lambda_i(x_j) = \begin{cases} x_j & i = 1 \\ 0 & i > 1, \end{cases}$$

for $j = 1, 2, \dots,$. Define a total ordering on compositions called the *well-ordering* (weight, length, lexicographic) by:

(1) If $|\alpha| > |\beta|$, then $\alpha >_{wll} \beta$.
(2) If $|\alpha| = |\beta|$ and $\ell(\alpha) > \ell(\beta)$, then $\alpha >_{wll} \beta$.
(3) If $|\alpha| = |\beta|$, $\ell(\alpha) = \ell(\beta)$, and $\alpha >_{lex} \beta$, then $\alpha >_{wll} \beta$.

For example,

$$523 >_{wll} 11213 >_{wll} 323 >_{wll} 143.$$

Hazewinkel proves [62] that applying λ_n to the monomial quasisymmetric function indexed by a Lyndon word α produces

$$\lambda_n(M_\alpha) = M_{\alpha^{\star n}} + (smaller),$$

where $\alpha^{\star n}$ denotes concatenation of α with itself n times and $(smaller)$ is a $\mathbb{Z}$-linear combination of monomial quasisymmetric functions which are wll-smaller than $\alpha^{\star n}$. For example,

$$\lambda_2(M_{(1,2)}) = M_{1212} + \text{ some subset of the set } V \cup W \cup Y,$$

where

$$V = \{\text{all words with weight } \le 5\},$$

$$W = \{\text{all words of weight 6 and length } \le 3\},$$

and

$$Y = \{1122, 1113, 1131\}.$$

Theorem 4.7. ([62]) *Let $eLYN$ be the set of all Lyndon words $u = u_1u_2\cdots u_m$ such that* $\gcd\{u_1, u_2, \ldots, u_m\} = 1$*. Then the set $\{\lambda_n(M_u)\}_{u \in eLYN}$ for all $n \in \mathbb{N}$ freely generates the ring of quasisymmetric functions over the integers.*

Although the monomial quasisymmetric functions are not multiplicative, Theorem 4.7 provides a way to construct a multiplicative generating set. Therefore, QSym is a polynomial algebra over the integers in the set $\{\lambda_n(M_\alpha)\}_{\alpha \in eLYN}$. Note that Theorem 4.7 therefore implies the Ditters conjecture.

5 Connections to Symmetric Functions and the Polynomial Ring

This section discusses several recent developments connecting quasisymmetric functions to important open problems within symmetric functions and the polynomial ring. We focus our scope to three topics: chromatic quasisymmetric functions, transitions from QSym to Sym, and liftings of QSym bases to the polynomial ring. We will unfortunately not be able to address the Eulerian quasisymmetric functions [113], which are in fact symmetric despite their definition in terms of quasisymmetric functions. See [113] for a wonderful introduction to these fascinating objects of study, including the important definitions and theorems as well as the research avenues they introduce. We also regretfully omit the recently developed theory of dual equivalence; see [7] for information about this new paradigm and how to use it.

5.1 Chromatic Quasisymmetric Functions

Let $G = (V, E)$ be a graph with vertices V and edges E and let S be a subset of the positive integers $\mathbb{P}$. A *proper S-coloring* of G is a function $\kappa\colon V \to S$ such that if two vertices i and j are adjacent (i.e., $\{i, j\} \in E$), then i and j are assigned different colors (i.e., $\kappa(i) \neq \kappa(j)$). The *chromatic number* $\chi(G)$ is the minimum number of colors (size of S) necessary to construct a proper S-coloring of G.

It is natural to ask how many proper $\{1, 2, \ldots, m\}$-colorings exist for a graph G; this number is denoted $\chi_G(m)$. It is a nonnegative integer when m is a positive integer, and it is a polynomial called the *chromatic polynomial* when m is an indeterminant. Stanley generalized this notion [118] to construct a symmetric function generated from the set $\mathcal{C}(G)$ of all proper $\mathbb{P}$-colorings of G as follows:

$$X_G(x) := \sum_{\kappa \in \mathcal{C}(G)} x_\kappa,$$

where $x = (x_1, x_2, \ldots)$ is a sequence of commuting variables and $x_\kappa = \prod_{v \in V} x_{\kappa(v)}$. Notice that plugging in $x_i = 1$ for all i produces the chromatic polynomial $X_G(1^m) = \chi_G(m)$. For example, the path P_3 on three vertices has chromatic symmetric function

$$X_{P_3}(x_1, x_2, x_3) = x_1^2 x_2 + x_1^2 x_3 + x_2^2 x_3 + x_1 x_2^2 + x_1 x_3^2 + x_2 x_3^2 + 6 x_1 x_2 x_3,$$

and $X_{P_3}(1, 1, 1) = 12$, the number of proper colorings of P_3 with 3 colors.

Recall that if a function has positive coefficients when expanded in a basis B, then it is said to be *B-positive*. For example, the elementary symmetric functions are Schur-positive since

$$e_\lambda = \sum_\mu K_{\mu'\lambda} s_\mu,$$

where $K_{\mu'\lambda}$ is the number of semi-standard Young tableaux of shape μ' and content λ. One significant open question about chromatic symmetric functions relates to positivity in the elementary basis for symmetric functions. If P is a partially ordered set, then the *incomparability graph* of P is the graph $inc(P)$ whose vertices are the elements of P and whose edges are the pairs of vertices which are incomparable in P. A poset is called *$(r + s)$-free* if no induced subposet is isomorphic to the direct sum of a chain (totally ordered set) with r elements and a chain with s elements.

Conjecture 5.1. (Stanley-Stembridge Conjecture [118, 121]) *If $G = inc(P)$ for some $(3 + 1)$-free poset P, then $X_G(x)$ is e-positive.*

For example, the poset P in Fig. 10 is $(3 + 1)$-free. The e-expansion for the chromatic symmetric function corresponding to its incomparability graph is $X_{inc(P)} = 4e_{31} + 8e_4$.

The incomparability graph for a $(3 + 1)$-free poset is an example of a *claw-free* graph. A claw-free graph is a graph which does not contain the star graph S_3 (depicted in Fig. 11) as a subgraph. However, not all claw-free graphs are e-positive; Dahlberg, Foley and van Willigenburg [27] provide a family of claw-free graphs which are not e-positive.

Gasharov [42] proved that the incomparability graph of a $(3 + 1)$-free poset is Schur-positive. Since the elementary symmetric functions are Schur-positive, Schur

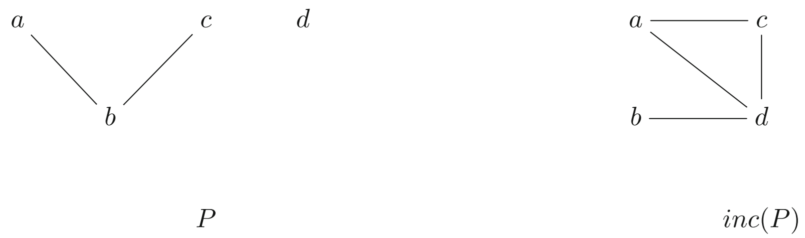


Fig. 10 A poset P and its incomparability graph

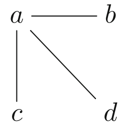


Fig. 11 Star graph S_3 on four vertices

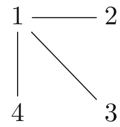


Fig. 12 A labeling of the star graph S_3 on four vertices

positivity would follow immediately from e-positivity. Guay-Paquet [48] proved that the chromatic symmetric function of a $(3+1)$-free poset is a convex combination of chromatic symmetric functions of posets which are both $(3+1)$-free and $(2+2)$-free. This reduces the e-positivity conjecture to a subclass of posets with more structure than posets which are $(3+1)$-free. A *natural unit interval order* is a poset P on the set $[n] := \{1, 2, \ldots, n\}$ obtained from a certain type of intervals on the real line as follows. Let $\{[a_1, a_1+1], [a_2, a_2+1], \ldots, [a_n, a_n+1]\}$ be a collection of closed intervals of length one such that $a_i < a_{i+1}$ for $1 \le i \le n-1$. Set $i <_P j$ if $a_i + 1 < a_j$. The resulting partially ordered set will always be $(3+1)$-free and $(2+2)$-free, and in fact, every poset that is both $(3+1)$-free and $(2+2)$-free is a unit interval order [112].

Shareshian and Wachs recently proposed a new approach to the Stanley–Stembridge e-positivity conjecture in the form of a refinement of Stanley's chromatic symmetric functions. This refinement behaves nicely with respect to unit interval orders.

Definition 5.2. ([115]) *Let $G = (V, E)$ be a graph whose vertex set V is a finite subset of $\mathbb{P}$. The* chromatic quasisymmetric function *of G is*

$$X_G(x,t) = \sum_{\kappa \in \mathcal{C}(G)} t^{asc(\kappa)} x^{\kappa},$$

where

$$asc(\kappa) = |\{\{i,j\} \in E \mid i < j \text{ and } \kappa(i) < \kappa(j)\}|.$$

Notice that the chromatic quasisymmetric function $X_G(x,t)$ depends not only on the isomorphism class of the graph G but also on the labeling of the vertices of G. Let G be the star graph on four vertices labeled as shown in Fig. 12. Then

$$X_G(x,t) = M_{31} + M_{121} + M_{211} + M_{1111} + t(M_{121} + 2M_{211} + 3M_{1111}) +$$

$$+ t^2(M_{121} + 2M_{211} + M_{112} + 5M_{1111}) + t^3(M_{121} + 2M_{112} + M_{211} + M_{13} + 6M_{1111}) +$$

$$+ t^4(M_{121} + 2M_{112} + 5M_{1111}) + t^5(M_{121} + M_{112} + 3M_{1111}) + t^6(M_{1111}).$$

The chromatic quasisymmetric function reduces to the chromatic symmetric function by setting $t = 1$; that is $X_G(x, 1) = X_G(x)$.

Let ω be the involution map on QSym which sends F_S to $F_{[n-1]\setminus S}$. The image of $X_G(x, t)$ under ω has a natural positive expansion into the fundamental basis for quasisymmetric functions [115]. Shareshian and Wachs [114] further conjecture that when G is the incomparability graph of a natural unit interval order, this image corresponds to a sum of Frobenius characteristics associated to certain Hessenberg varieties. This conjecture was proved by Brosnan and Chow [22] and, through a different approach, by Guay-Paquet [49], providing an alternate proof of Schur positivity.

Theorem 5.3. ([115]) *If G is the incomparability graph of a natural unit interval order, then $X_G(x, t)$ is symmetric in the x-variables.*

Not every graph whose chromatic quasisymmetric function is symmetric is an incomparability graph of a natural unit interval order. One interesting open question is to classify which graphs admit a symmetric chromatic quasisymmetric function.

Several extensions of chromatic quasisymmetric functions have recently emerged, demonstrating the many different areas this research impacts. Ellzey extends this paradigm to directed graphs [35]. Haglund and Wilson express the integral form Macdonald polynomials as weighted sums of chromatic quasisymmetric functions [53]. Clearman, Hyatt, Shelton, and Skandera interpret the chromatic quasisymmetric functions in terms of Hecke algebra traces [26], while Alexandersson and Panova connect the chromatic quasisymmetric functions to LLT polynomials [2].

5.2 *Quasisymmetric Expansions of Symmetric Functions*

As quasisymmetric functions become more ubiquitous, many natural expansions of symmetric functions into quasisymmetric functions (particularly into the fundamental quasisymmetric functions) are appearing. It is natural to try to use this structure to answer classical questions about symmetric functions such as Schur positivity. Egge, Loehr, and Warrington [33] recently introduced a method to convert the quasisymmetric expansion of a symmetric function into the Schur function expansion, providing a new approach to questions of Schur positivity.

We need several definitions in order to describe the "modified inverse Kostka matrix" and some interesting applications of this paradigm. A *rim-hook* is a set of contiguous cells in a partition diagram such that each diagonal contains at most one cell. A *special rim-hook tableau* is a decomposition of a partition diagram into rim-hooks such that each rim-hook contains at least one cell in the leftmost column of the diagram. Eğecioğlu and Remmel [36] use special rim-hook tableaux in their formula for the *inverse Kostka matrix*, which is the transition matrix from the monomial basis for symmetric functions to the Schur functions.

The *sign* of a special rim-hook is $(-1)^{r-1}$, where r is the number of rows spanned by the rim-hook. The sign of a special rim-hook tableau is the product of the signs of

its rim-hooks. A special rim-hook tableau is said to be *flat* if each rim-hook contains exactly one cell in the leftmost column of the partition diagram.

Theorem 5.4. ([33]) *Let F be a field, and let f be a symmetric function given by its expansion into the fundamental quasisymmetric functions so that*

$$f = \sum_{\alpha \models n} y_\alpha F_\alpha.$$

Then the coefficients x_λ in the Schur function expansion $f = \sum_{\lambda \vdash n} x_\lambda s_\lambda$ are given by

$$x_\lambda = \sum_{\alpha \models n} y_\alpha K_n^*(\alpha, \lambda),$$

where $K_n^*(\alpha, \lambda)$ is the sum of the signs of all flat special rim-hook tableaux of partition shape λ and content α.

Theorem 5.4 provides a potential alternative approach to proving that Macdonald polynomials expand positively into the Schur functions. In particular, recall that Theorem 3.2 describes a formula for expanding Macdonald polynomials into the fundamental quasisymmetric functions. Combining this formula with Theorem 5.4 implies that the coefficient of s_λ in the Schur function expansion of $\tilde{H}_\mu$ is given by

$$\sum_{\alpha \vdash n} K_n^*(\alpha, \lambda) \left(\sum_{\beta \in \{\mathfrak{S}_n \mid \mathrm{Des}(\beta^{-1}) = \alpha\}} q^{\mathrm{inv}(\beta,\mu)} t^{\mathrm{maj}(\beta,\mu)} \right).$$

The following example is similar to that appearing in [33]. If $\mu = (3, 1)$ and $\lambda = (2, 2)$, then there exists a flat special rim-hook tableau for $\alpha = (2, 2)$ and a flat special rim-hook tableau $\alpha = (1, 3)$. These correspond to

$$K_4^*((2, 2), (2, 2)) = +1 \text{ and } K_4^*((1, 3), (2, 2)) = -1$$

respectively. The permutations w whose inverse descent sets $\mathrm{Des}(w^{-1})$ are $\{2\}$ are 3412, 3142, 3124, 1324, and 1342. Computing the inv and maj for the fillings of $(3, 1)$ with these permutations as reading words produces $2q^2 + qt + t + q$. Similarly, the permutations whose inverse descent sets are $\{1\}$ are 2341, 2314, and 2134. Their inv and maj (for fillings of $(3, 1)$) produce $-q^2 - q - t$. Putting this together, the coefficient of s_{22} in H_{31} is

$$2q^2 + qt + t + q - (q^2 + q + t) = q^2 + qt.$$

Notice that negative terms do appear in the $K_n^*(\alpha, \lambda)$. This means that in order to apply this technique to the Schur positivity of Macdonald polynomials problem, one must find involutions to cancel out the negative terms.

A further application of this transition matrix from the fundamental quasisymmetric functions to Schur functions is to the Foulkes Plethysm Conjecture [38], which states that $s_n[s_m] - s_m[s_n]$ (where the brackets denote a certain type of substitution called *plethysm*) is Schur-positive. Loehr and Warrington [83] provide a formula for the expansion of $s_\mu[s_\nu]$ into fundamental quasisymmetric functions using a novel interpretation of the "reading word" of a matrix. The modified inverse Kostka matrix could then be used to determine the Schur function expansion of $s_n[s_m] - s_m[s_n]$, again with the caveat that involutions are needed to cancel out the negative terms.

Garsia and Remmel recently found a further extension of the Egge, Loehr, Warrington result. They proved that each fundamental appearing in the fundamental expansion of a symmetric function can be replaced by the Schur function indexed by the same composition. Since every such Schur function is either 0 or $\pm s_\lambda$ for some partition λ, this expansion can be simplified to a signed sum of Schur functions indexed by partitions.

Theorem 5.5. ([40]) *Let f be a symmetric function which is homogeneous of degree n and expands into the fundamental basis for quasisymmetric functions as follows:*

$$f = \sum_{\alpha \models n} a_\alpha F_\alpha.$$

Then

$$f = \sum_{\alpha \models n} a_\alpha s_\alpha.$$

Theorem 5.5 already has a number of important consequences. Garsia and Remmel used this approach to formulate a conjecture regarding the modified Hall–Littlewood polynomials. Leven applied this method to prove an extension of the Shuffle Conjecture for the cases $m = 2$ and $n = 2$ [80]. Qiu and Remmel considered the cases of this "Rational Shuffle Conjecture" where m or n equals 3 [102].

5.3 Slide Polynomials and the Quasi-key Basis

Schubert polynomials are an important class of polynomials, first introduced by Lascoux and Schützenberger [77] to provide a new method for computing intersection numbers in the cohomology ring of the complete flag variety. Several different combinatorial formulas for Schubert polynomials have been discovered since their original introduction as divided difference operators, including but not limited to reduced decompositions [21, 37] and RC-graphs [15]. Despite the numerous ways to construct Schubert polynomials, it remains an open problem to provide a combinatorial formula for the expansion of a product of Schubert polynomials into the Schubert basis.

Assaf and Searles [5] further the study of Schubert polynomials with the introduction of two new families of polynomials, both of which positively refine the Schubert polynomials. These new families, called the *monomial slide polynomials* and the *fundamental slide polynomials*, exhibit positive structure constants (meaning the coefficients appearing in their products are always positive), whereas the key polynomials (another family of polynomials refining the Schubert polynomials [28, 78, 105]) have signed structure constants. Although the slide polynomials have many interesting applications (to Schubert polynomials and other objects of study in algebraic combinatorics), this article focuses on their connections to quasisymmetric functions.

Remove the zeros from a weak composition γ to obtain a (strong) composition called the *flattening* of γ, denoted $\flat(\gamma)$. The *monomial slide polynomial* $\mathfrak{M}_\gamma$ is then defined by

$$\mathfrak{M}_\gamma(x_1, x_2, \ldots, x_n) = \sum_{\substack{\delta \geq \gamma \\ \flat(\delta)=\flat(\gamma)}} x_1^{\delta_1} x_2^{\delta_2} \cdots x_n^{\delta_n},$$

where $\delta \geq \gamma$ if δ *dominates* γ; that is $\delta_1 + \delta_2 + \cdots + \delta_i \geq \gamma_1 + \gamma_2 + \cdots + \gamma_i$ for all $1 \leq i \leq n$. The related *fundamental slide polynomial* $\mathfrak{F}_\gamma$ is defined by

$$\mathfrak{F}_\gamma = \sum_{\substack{\delta \geq \gamma \\ \flat(\delta) \text{ refines } \flat(\gamma)}} x_1^{\delta_1} x_2^{\delta_2} \cdots x_n^{\delta_n}.$$

For example,

$$\begin{aligned}\mathfrak{F}_{1032}(x_1, x_2, x_3, x_4) &= x_1x_3^3x_4^2 + x_1x_2^3x_4^2 + x_1x_2^3x_3^2 + x_1x_2x_3^2x_4^2 + x_1x_2^2x_3x_4^2 + x_1x_2^3x_3x_4 \\ &= \mathfrak{M}_{1032} + \mathfrak{M}_{1122} + \mathfrak{M}_{1212} + \mathfrak{M}_{1311}.\end{aligned}$$

Each of these families $\{\mathfrak{M}_\gamma\}_\gamma$ and $\{\mathfrak{F}_\gamma\}_\gamma$ (indexed by weak compositions of k) of polynomials is a $\mathbb{Z}$-basis for polynomials of degree k in n variables.

Assaf and Searles [4] also introduce a related basis, called the *quasi-key polynomials* $\mathfrak{Q}_\gamma$, for the polynomial ring which is analogous to the key polynomials. These polynomials are positive sums of fundamental slide polynomials and in fact stabilize to the quasisymmetric Schur functions as zeros are prepended to their indexing compositions. (Prepending m zeros to the composition γ is denoted by $0^m \times \gamma$.)

Theorem 5.6. ([4]) *For any weak composition γ, we have*

$$\lim_{m\to\infty} \mathfrak{Q}_{0^m\times\gamma} = \mathcal{S}_{\flat(\gamma)}.$$

Each Schubert polynomial can be written as a positive sum of fundamental slide polynomials using a new object called a *quasi-Yamanouchi pipe dream*. While this definition takes us too far from our current topic, we do take the time to describe a closely related construction involving the fundamental expansion of Schur functions.

Recall that the Schur functions decompose into a positive sum of the fundamental quasisymmetric functions (see Eq. 1.1); this formula can be computed by finding the descent sets of all standard Young tableaux of a given shape. However, when the number of variables is less than the number of descents, the corresponding fundamental equals 0. Assaf and Searles [5] introduce a class of semi-standard Young tableaux, called *quasi-Yamanouchi tableaux*, which dictate precisely which fundamentals appear in the decomposition with nonzero coefficient when the variables are restricted.

Definition 5.7. ([5]) *A semi-standard Young tableau is said to be* quasi-Yamanouchi *if for all* $i > 1$, *the leftmost occurrence of* i *lies weakly left of some appearance of* $i - 1$. *Let* $QYT_n(\lambda)$ *denote the set of quasi-Yamanouchi tableaux of shape* λ *whose entries are in* $[n]$.

The *weight* of a quasi-Yamanouchi tableau T is given by $wt(T) = \prod_i x_i^{m_i}$, where m_i is the number of times the entry i appears in T.

Theorem 5.8. ([5]) *The Schur polynomial* $s_\lambda(x_1, \ldots, x_n)$ *is given by*

$$s_\lambda(x_1, \ldots, x_n) = \sum_{T \in QYT_n(\lambda)} F_{wt(T)}(x_1, \ldots, x_n).$$

For example, the three quasi-Yamanouchi tableaux of shape $\lambda = (4, 2)$ and entries in $\{1, 2\}$ are

$$\begin{array}{|c|c|cc} \hline 2 & 2 & \multicolumn{2}{c}{} \\ \hline 1 & 1 & \multicolumn{1}{|c|}{1} & \multicolumn{1}{c|}{1} \\ \hline \end{array}, \quad \begin{array}{|c|c|cc} \hline 2 & 2 & \multicolumn{2}{c}{} \\ \hline 1 & 1 & \multicolumn{1}{|c|}{1} & \multicolumn{1}{c|}{2} \\ \hline \end{array}, \quad \begin{array}{|c|c|cc} \hline 2 & 2 & \multicolumn{2}{c}{} \\ \hline 1 & 1 & \multicolumn{1}{|c|}{2} & \multicolumn{1}{c|}{2} \\ \hline \end{array},$$

and therefore, the Schur expansion into fundamentals is

$$s_{42}(x_1, x_2) = F_{4,2}(x_1, x_2) + F_{3,3}(x_1, x_2) + F_{2,4}(x_1, x_2).$$

Note that all the terms appearing on the right-hand side are nonzero, and there is no need to calculate the descent sets for all nine standard Young tableaux of shape (4, 2). This is important in the study of Schubert polynomials because although certain classes of Schubert polynomials are equal to Schur functions, the number of variables appearing varies based on the indexing permutation. For example, the Schubert polynomial indexed by the permutation 213 (written in one-line notation) is equal to the Schur function $s_1(x_1) = x_1$ while the Schubert polynomial indexed by the permutation 132 (written in one-line notation) is $s_1(x_1, x_2) = x_1 + x_2$. A thorough understanding of precisely the nonzero terms appearing in the quasisymmetric expansion is therefore crucial to the quest of proving a combinatorial formula for Schubert multiplication.

This connection to Schubert multiplication (a long-standing open problem in algebraic combinatorics) exemplifies the utility of quasisymmetric functions. Quasisymmetric functions appear in a number of other important problems which have helped

to shape the study of algebraic combinatorics including Schur positivity of Macdonald polynomials, the Foulkes plethysm conjecture, and the Stanley-Stembridge conjecture. We hope the reader comes away from this article with a deeper appreciation for the beauty and utility of quasisymmetric functions and a desire to further explore this exciting and far-reaching avenue of research.

Acknowledgements I am very grateful to Hélène Barcelo, Gizem Karaali, and Rosa Orellana for inviting me to produce this chapter. I would also like to thank Ed Allen, Susanna Fishel, Josh Hallam, Jim Haglund, and John Shareshian for helpful feedback along the way. Finally, I greatly appreciate the insightful comments from a diligent anonymous referee.

References

1. M. Aguiar, N. Bergeron, F. Sottile, Combinatorial Hopf algebras and generalized Dehn-Sommerville relations. Compos. Math. **142**(1), 1–30 (2006)
2. P. Alexandersson, G. Panova, LLT polynomials, chromatic quasisymmetric functions and graphs with cycles. arXiv:1705.10353
3. E.E. Allen, J. Hallam, S.K. Mason, Dual immaculate quasisymmetric functions expand positively into Young quasisymmetric Schur functions (2016), arXiv:1606.03519
4. S. Assaf, D. Searles, Kohnert tableaux and a lifting of quasi-Schur functions (2016), arXiv:1609.03507
5. S. Assaf, D. Searles, Schubert polynomials, slide polynomials, Stanley symmetric functions and quasi-Yamanouchi pipe dreams. Adv. Math. **306**, 89–122 (2017)
6. S.H. Assaf, Dual equivalence graphs, ribbon tableaux and Macdonald polynomials. ProQuest LLC, Ann Arbor, MI, 2007. Thesis (Ph.D.)-University of California, Berkeley (2007)
7. S.H. Assaf, Dual equivalence graphs I: a new paradigm for Schur positivity. Forum Math. Sigma **3: e12**, 33 (2015)
8. J.-C. Aval, N. Bergeron, Catalan paths and quasi-symmetric functions. Proc. Am. Math. Soc. **131**(4), 1053–1062 (2003)
9. J.-C. Aval, F. Bergeron, N. Bergeron, Ideals of quasi-symmetric functions and super-covariant polynomials for $\mathscr{S}_n$. Adv. Math. **181**(2), 353–367 (2004)
10. A. Baker, B. Richter, Quasisymmetric functions from a topological point of view. Math. Scand. **103**(2), 208–242 (2008)
11. C. Ballantine, Z. Daugherty, A. Hicks, S. Mason, E. Niese, Quasisymmetric power sums (2017), arXiv:1710.11613
12. C. Berg, N. Bergeron, F. Saliola, L. Serrano, M. Zabrocki, A lift of the Schur and Hall-Littlewood bases to non-commutative symmetric functions. Can. J. Math. **66**(3), 525–565 (2014)
13. C. Berg, N. Bergeron, F. Saliola, L. Serrano, M. Zabrocki, Indecomposable modules for the dual immaculate basis of quasi-symmetric functions. Proc. Am. Math. Soc. **143**(3), 991–1000 (2015)
14. F. Bergeron, *Algebraic Combinatorics and Coinvariant Spaces*. CMS Treatises in Mathematics. Canadian Mathematical Society, Ottawa, ON; A K Peters, Ltd., Wellesley, MA, 2009
15. N. Bergeron, S. Billey, RC-graphs and Schubert polynomials. Exp. Math. **2**(4), 257–269 (1993)
16. N. Bergeron, S. Mykytiuk, F. Sottile, S. van Willigenburg, Shifted quasi-symmetric functions and the Hopf algebra of peak functions. Discret. Math. **246**(1–3), 57–66 (2002) (Formal power series and algebraic combinatorics (Barcelona, 1999))
17. N. Bergeron, S. Mykytiuk, F. Sottile, S. van Willigenburg, Noncommutative Pieri operators on posets. J. Comb. Theory Ser. A **91**(1–2), 84–110 (2000) (In memory of Gian-Carlo Rota)

18. N. Bergeron, J. Sánchez-Ortega, M. Zabrocki, The Pieri rule for dual immaculate quasi-symmetric functions. Ann. Comb. **20**(2), 283–300 (2016)
19. L.J. Billera, N. Liu, Noncommutative enumeration in graded posets. J. Algebr. Comb. **12**(1), 7–24 (2000)
20. L.J. Billera, S.K. Hsiao, S. van Willigenburg, Peak quasisymmetric functions and Eulerian enumeration. Adv. Math. **176**(2), 248–276 (2003)
21. S.C. Billey, W. Jockusch, R.P. Stanley, Some combinatorial properties of Schubert polynomials. J. Algebr. Comb. **2**(4), 345–374 (1993)
22. P. Brosnan, T.Y. Chow, Unit interval orders and the dot action on the cohomology of regular semisimple Hessenberg varieties (2015), arXiv:1511.00773
23. R.W. Carter, Representation theory of the 0-Hecke algebra. J. Algebra **104**(1), 89–103 (1986)
24. I. Cherednik, Nonsymmetric Macdonald polynomials. Int. Math. Res. Not. **10**, 483–515 (1995)
25. C.-O. Chow, Noncommutative symmetric functions of type B. ProQuest LLC, Ann Arbor, MI, 2001. Thesis (Ph.D.)–Massachusetts Institute of Technology
26. S. Clearman, M. Hyatt, B. Shelton, M. Skandera, Evaluations of Hecke algebra traces at Kazhdan-Lusztig basis elements. Electron. J. Comb. **23**(2), Paper 2.7, 56 (2016)
27. S. Dahlberg, A. Foley, S. van Willigenburg, Resolving Stanley's e-positivity of claw-contractible-free graphs (2017), arXiv:1703.05770
28. M. Demazure, Une nouvelle formule des caractères. Bull. Sci. Math. (2) **98**(3), 163–172 (1974)
29. H. Derksen, Symmetric and quasi-symmetric functions associated to polymatroids. J. Algebr. Comb. **30**(1), 43–86 (2009)
30. E.J. Ditters, Curves and formal (co)groups. Invent. Math. **17**, 1–20 (1972)
31. G. Duchamp, D. Krob, B. Leclerc, J.-Y. Thibon, Fonctions quasi-symétriques, fonctions symétriques non commutatives et algèbres de Hecke à $q = 0$. C. R. Acad. Sci. Paris Sér. I Math. **322**(2), 107–112 (1996)
32. G. Duchamp, F. Hivert, J.-Y. Thibon, Noncommutative symmetric functions. VI. Free quasi-symmetric functions and related algebras. Int. J. Algebra Comput. **12**(5), 671–717 (2002)
33. E. Egge, N.A. Loehr, G.S. Warrington, From quasisymmetric expansions to Schur expansions via a modified inverse Kostka matrix. Eur. J. Comb. **31**(8), 2014–2027 (2010)
34. R. Ehrenborg, On posets and Hopf algebras. Adv. Math. **119**(1), 1–25 (1996)
35. B. Ellzey, A directed graph generalization of chromatic quasisymmetric functions (2017), arXiv:1709.00454
36. Ö. Eğecioğlu, J.B. Remmel, A combinatorial interpretation of the inverse Kostka matrix. Linear Multilinear Algebr. **26**(1–2), 59–84 (1990)
37. S. Fomin, R. Stanley, Schubert polynomials and the nil-Coxeter algebra. Adv. Math. **103**(2), 196–207 (1994)
38. H.O. Foulkes, Concomitants of the quintic and sextic up to degree four in the coefficients of the ground form. J. Lond. Math. Soc. **25**, 205–209 (1950)
39. W. Fulton, *Young Tableaux*, vol. 35. London Mathematical Society Student Texts (Cambridge University Press, Cambridge, 1997)
40. A. Garsia, J. Remmel, A note on passing from a quasi-symmetric function expansion to a Schur function expansion of a symmetric function (2018), arXiv:1802.09686
41. A.M. Garsia, N. Wallach, Qsym over Sym is free. J. Comb. Theory Ser. A **104**(2), 217–263 (2003)
42. V. Gasharov, Incomparability graphs of $(3+1)$-free posets are s-positive, in *Proceedings of the 6th Conference on Formal Power Series and Algebraic Combinatorics* (New Brunswick, NJ, 1994), vol. 157 (1996), pp. 193–197
43. I.M. Gelfand, D. Krob, A. Lascoux, B. Leclerc, V.S. Retakh, J.-Y. Thibon, Noncommutative symmetric functions. Adv. Math. **112**(2), 218–348 (1995)
44. I.M. Gessel, Multipartite P-partitions and inner products of skew Schur functions. Contemp. Math. **34**, 289–301 (1984)
45. I.M. Gessel, A historical survey of P-partitions, *The Mathematical Legacy of Richard P. Stanley* (American Mathematical Society, Providence, 2016), pp. 169–188

46. D. Grinberg, Dual creation operators and a dendriform algebra structure on the quasisymmetric functions. Can. J. Math. **69**(1), 21–53 (2017)
47. D. Grinberg, V. Reiner, Hopf algebras in combinatorics (2014), arXiv:1409.8356
48. M. Guay-Paquet, A modular relation for the chromatic symmetric functions of (3+1)-free posets (2013), arXiv:1306.2400
49. M. Guay-Paquet, A second proof of the Shareshian–Wachs conjecture, by way of a new Hopf algebra (2016), arXiv:1601.05498
50. J. Haglund, A combinatorial model for the Macdonald polynomials. Proc. Natl. Acad. Sci. USA **101**(46), 16127 (2004)
51. J. Haglund, The genesis of the Macdonald polynomial statistics. Sém. Lothar. Comb. **54A**, Art. B54Ao, 16 (2005/07)
52. J. Haglund, *The q,t-Catalan Numbers and the Space of Diagonal Harmonics*, vol. 41. University Lecture Series (American Mathematical Society, Providence, 2008)
53. J. Haglund, A.T. Wilson, Macdonald polynomials and chromatic quasisymmetric functions (2017)
54. J. Haglund, M. Haiman, N. Loehr, A combinatorial formula for Macdonald polynomials. J. Am. Math. Soc. **18**, 735–761 (2004)
55. J. Haglund, M. Haiman, N. Loehr, Combinatorial theory of Macdonald polynomials I: proof of Haglund's formula. Proc. Natl. Acad. Sci. **102**(8), 2690 (2005)
56. J. Haglund, M. Haiman, N. Loehr, A combinatorial formula for nonsymmetric Macdonald polynomials. Am. J. Math. **130**(2), 359–383 (2008)
57. J. Haglund, K. Luoto, S. Mason, S. van Willigenburg, Quasisymmetric Schur functions. J. Comb. Theory Ser. A **118**(2), 463–490 (2011)
58. J. Haglund, K. Luoto, S. Mason, S. van Willigenburg, Refinements of the Littlewood-Richardson rule. Trans. Am. Math. Soc. **363**(3), 1665–1686 (2011)
59. M. Haiman, Hilbert, schemes, polygraphs and the Macdonald positivity conjecture. J. Am. Math. Soc. **14**(4), 941–1006 (electronic) (2001)
60. M. Hazewinkel, The algebra of quasi-symmetric functions is free over the integers. Adv. Math. **164**(2), 283–300 (2001)
61. M. Hazewinkel, Symmetric functions, noncommutative symmetric functions, and quasisymmetric functions. Acta Appl. Math. **75**(1–3), 55–83 (2003) (Monodromy and differential equations (Moscow, 2001))
62. M. Hazewinkel, Explicit polynomial generators for the ring of quasisymmetric functions over the integers. Acta Appl. Math. **109**(1), 39–44 (2010)
63. M. Hazewinkel, N. Gubareni, V.V. Kirichenko, *Algebras, Rings and Modules*. Volume 168 of *Mathematical Surveys and Monographs* (American Mathematical Society, Providence, 2010) (Lie algebras and Hopf algebras)
64. P. Hersh, S.K. Hsiao, Random walks on quasisymmetric functions. Adv. Math. **222**(3), 782–808 (2009)
65. F. Hivert, Analogues non-commutatifs et quasi-symétriques des fonctions de Hall-Littlewood, et modules de Demazure d'une algèbre enveloppante quantique dégénérée. C. R. Acad. Sci. Paris Sér. I Math. **326**(1), 1–6 (1998)
66. F. Hivert, Hecke algebras, difference operators, and quasi-symmetric functions. Adv. Math. **155**(2), 181–238 (2000)
67. M.E. Hoffman, Quasi-symmetric functions and mod p multiple harmonic sums. Kyushu J. Math. **69**(2), 345–366 (2015)
68. S.K. Hsiao, T.K. Petersen, The Hopf algebras of type B quasisymmetric functions and peak functions (2006), arXiv:math/0610976
69. S.K. Hsiao, G. Karaali, Multigraded combinatorial Hopf algebras and refinements of odd and even subalgebras. J. Algebr. Comb. **34**(3), 451–506 (2011)
70. S.K. Hsiao, T. Kyle Petersen, Colored posets and colored quasisymmetric functions. Ann. Comb. **14**(2), 251–289 (2010)
71. J. Huang, 0-Hecke algebra actions on coinvariants and flags. J. Algebr. Comb. **40**(1), 245–278 (2014)

72. J. Huang, 0-Hecke algebra action on the Stanley-Reisner ring of the Boolean algebra. Ann. Comb. **19**(2), 293–323 (2015)
73. J. Huang, A tableau approach to the representation theory of 0-Hecke algebras. Ann. Comb. **20**(4), 831–868 (2016)
74. S.A. Joni, G.-C. Rota, Coalgebras and bialgebras in combinatorics. Stud. Appl. Math. **61**(2), 93–139 (1979)
75. D.E. Knuth, Permutations, matrices, and generalized young tableaux. Pac. J. Math. **34**(3), 709–727 (1970)
76. D. Krob, J.-Y. Thibon, Noncommutative symmetric functions. IV. Quantum linear groups and Hecke algebras at $q = 0$. J. Algebr. Comb. **6**(4), 339–376 (1997)
77. A. Lascoux, M.P. Schützenberger, Polynômes de Schubert. C. R. Acad. Sci. Paris Sér. I Math. **294**(13), 447–450 (1982)
78. A. Lascoux, M.-P. Schützenberger, Keys & standard bases, in *Invariant Theory and Tableaux (Minneapolis, MN, 1988)*. Volume 19 of IMA Volumes in Mathematics and its Applications (Springer, New York, 1990), pp. 125–144
79. A. Lauve, S.K. Mason, QSym over Sym has a stable basis. J. Comb. Theory Ser. B **118**(5), 1661–1673 (2011)
80. E. Leven, Two special cases of the rational shuffle conjecture, in *26th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2014), Discrete Mathematics & Theoretical Computer Science Proceedings, AT*. (Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2014), pp. 789–800
81. Y. Li, Toward a polynomial basis of the algebra of peak quasisymmetric functions. J. Algebr. Comb. **44**(4), 931–946 (2016)
82. D.E. Littlewood, On certain symmetric functions. Proc. Lond. Math. Soc. **3**(11), 485–498 (1961)
83. N.A. Loehr, G.S. Warrington, Quasisymmetric expansions of Schur-function plethysms. Proc. Am. Math. Soc. **140**(4), 1159–1171 (2012)
84. M. Lothaire, *Combinatorics on Words*. Cambridge Mathematical Library (Cambridge University Press, Cambridge, 1997)
85. K. Luoto, S. Mykytiuk, S. van Willigenburg, *An Introduction to Quasisymmetric Schur Functions*. Springer Briefs in Mathematics (Springer, New York, 2013)
86. I.G. Macdonald, A new class of symmetric functions. Sémin. Lothar. Comb. 20 (1988)
87. I.G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd edn. Oxford Mathematical Monographs (The Clarendon Press, Oxford University Press, New York, 1995)
88. I.G. Macdonald, *Symmetric Functions and Orthogonal Polynomials*, vol. 12. University Lecture Series (American Mathematical Society, Providence, 1998)
89. C. Malvenuto, Produits et coproduits des fonctions quasi-symétriques et de l'alg‘ebre des descentes. PhD thesis, Laboratoire de Combinatoire et d'Informatique Mathématique UQAM (1994)
90. C. Malvenuto, C. Reutenauer, Duality between quasi-symmetric functions and the Solomon descent algebra. J. Algebr. **177**(3), 967–982 (1995)
91. S. Mason, A decomposition of Schur functions and an analogue of the Robinson-Schensted-Knuth algorithm. Sémin. Lothar. Comb. **57**, B57e (2008)
92. S. Mason, An explicit construction of type A Demazure atoms. J. Algebr. Comb. **29**(3), 295–313 (2009)
93. P.-L. Méliot, *Representation Theory of Symmetric Groups*. Discrete Mathematics and its Applications (Boca Raton) (CRC Press, Boca Raton, 2017)
94. F.D. Murnaghan, On the representations of the symmetric group. Am. J. Math. **59**(3), 437–488 (1937)
95. T. Nakayama, On some modular properties of irreducible representations of a symmetric group I. Jpn. J. Math. **18**, 89–108 (1941)
96. P.N. Norton, 0-Hecke algebras. J. Austral. Math. Soc. Ser. A **27**(3), 337–357 (1979)
97. J.-C. Novelli, J.-Y. Thibon, L.K. Williams, Combinatorial Hopf algebras, noncommutative Hall-Littlewood functions, and permutation tableaux. Adv. Math. **224**(4), 1311–1348 (2010)

98. J.-C. Novelli, L. Tevlin, J.-Y. Thibon, On some noncommutative symmetric functions analogous to Hall-Littlewood and Macdonald polynomials. Int. J. Algebr. Comput. **23**(4), 779–801 (2013)
99. E.M. Opdam, Harmonic analysis for certain representations of graded Hecke algebras. Acta Math. **175**(1), 75–121 (1995)
100. T. Kyle Petersen, A note on three types of quasisymmetric functions. Electron. J. Comb. **12**, Research Paper 61, 10 (2005)
101. T. Kyle Petersen, Enriched, P-partitions and peak algebras. Adv. Math. **209**(2), 561–610 (2007)
102. D. Qiu, J. Remmel, Schur function expansions and the rational shuffle conjecture. Sém. Lothar. Comb. **78B**, Art. 83, 13 (2017)
103. D.E. Radford, A natural ring basis for the shuffle algebra and an application to group schemes. J. Algebra **58**(2), 432–454 (1979)
104. A. Ram, M. Yip, A combinatorial formula for Macdonald polynomials. Adv. Math. **226**(1), 309–331 (2011)
105. V. Reiner, M. Shimozono, Key polynomials and a flagged Littlewood-Richardson rule. J. Comb. Theory Ser. A **70**(1), 107–143 (1995)
106. J.B. Remmel, R. Whitney, Multiplying Schur functions. J. Algorithms **5**(4), 471–487 (1984)
107. C. Reutenauer, *Free Lie Algebras*, vol. 7. London Mathematical Society Monographs. New Series (The Clarendon Press, Oxford University Press, New York, 1993)
108. B.E. Sagan, *The Symmetric Group*. Volume 203 of *Graduate Texts in Mathematics*, 2nd edn. (Springer, New York, 2001)
109. S. Sahi, Nonsymmetric, Koornwinder polynomials and duality. Ann. Math. (2) **150**(1), 267–282 (1999)
110. C. Schensted, Longest increasing and decreasing subsequences. Can. J. Math. **13**, 179–191 (1961)
111. W.R. Schmitt, Incidence Hopf algebras. J. Pure Appl. Algebra **96**(3), 299–330 (1994)
112. D. Scott, P. Suppes, Foundational aspects of theories of measurement. J. Symb. Log. **23**, 113–128 (1958)
113. J. Shareshian, M.L. Wachs, Eulerian quasisymmetric functions. Adv. Math. **225**(6), 2921–2966 (2010)
114. J. Shareshian, M.L. Wachs, Chromatic quasisymmetric functions and Hessenberg varieties, in *Configuration Spaces*. Volume 14 of CRM Series. Ed. Norm. (Pisa, 2012), pp. 433–460
115. J. Shareshian, M.L. Wachs, Chromatic quasisymmetric functions. Adv. Math. **295**, 497–551 (2016)
116. L. Solomon, A Mackey formula in the group ring of a Coxeter group. J. Algebra **41**(2), 255–264 (1976)
117. R.P. Stanley, *Ordered Structures and Partitions*. Memoirs of the American Mathematical Society, No. 119 (American Mathematical Society, Providence, 1972)
118. R.P. Stanley, A symmetric function generalization of the chromatic polynomial of a graph. Adv. Math. **111**(1), 166–194 (1995)
119. R.P. Stanley, Enumerative combinatorics, *62 of Cambridge Studies in Advanced Mathematics*, vol. 2 (Cambridge University Press, Cambridge, 1999)
120. R.P. Stanley, Generalized riffle shuffles and quasisymmetric functions. Ann. Comb. **5**(3–4), 479–491 (2001)
121. R.P. Stanley, J.R. Stembridge, On immanants of Jacobi-Trudi matrices and permutations with restricted position. J. Comb. Theory Ser. A **62**(2), 261–279 (1993)
122. J.R. Stembridge, Enriched P-partitions. Trans. Am. Math. Soc. **349**(2), 763–788 (1997)
123. V.V. Tewari, S.J. van Willigenburg, Modules of the 0-Hecke algebra and quasisymmetric Schur functions. Adv. Math. **285**, 1025–1065 (2015)
124. M. Yip, A Littlewood-Richardson rule for Macdonald polynomials. Math. Z. **272**(3–4), 1259–1290 (2012)
125. A.V. Zelevinsky, *Representations of Finite Classical Groups*, vol. 869. Lecture Notes in Mathematics (Springer, Berlin, 1981)

On Standard Young Tableaux of Bounded Height

M. J. Mishna

Abstract We survey some recent works on standard Young tableaux of bounded height. We focus on consequences resulting from numerous bijections to lattice walks in Weyl chambers.

1 Introduction

Standard Young tableaux are a classic object of mathematics, appearing in problems from representation theory to bijective combinatorics. Lattice walks restricted to cones are similarly a fundamental family, and they encode a wide variety of combinatorial structures from formal languages to queues. Standard Young tableaux of bounded height are in bijection with several different straightforward classes of lattice walks. This connection not only elucidates several sources of ubiquity on both accounts, but facilitates exact and asymptotic enumeration, as well as parameter analysis. This survey describes the cross developments over the past 30 years, and highlights some open problems. For more background on the tableaux, we recommend the surveys [1, 30] and particularly the article of Stanley [34].

We begin by fixing our notation and conventions. Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_t)$ be a partition of n into t parts. We write $\lambda \vdash n$ and $\ell(\lambda) = t$ in this case. The *Ferrers diagram of shape* λ is a representation of λ comprised of boxes indexed by pairs $\{(i, j) : 1 \leq i \leq t; 1 \leq j \leq \lambda_i\}$. Such a diagram is of size n and height t. A key parameter in our study is the number of columns of odd length.

A *standard Young tableau* of size n is a filling of a Ferrers diagram using precisely the integers 1 to n. The entries strictly increase to the right along each row and strictly increase down each column. A tableau is *semi-standard* if the entries weakly increase along each row and strictly increase down each column. In this work, we are interested

M. J. Mishna (✉)
Department of Mathematics, Simon Fraser University, 8888 University Drive,
Burnaby, BC V5A 1S5, Canada
e-mail: mmishna@sfu.ca

H. Barcelo et al. (eds.), *Recent Trends in Algebraic Combinatorics*,
Association for Women in Mathematics Series 16,
https://doi.org/10.1007/978-3-030-05141-9_8

in the number of standard Young tableaux of a given size with the height is bounded by a fixed value.

1.1 Enumeration Formulas

Rather classically, the number of standard Young tableaux of shape λ is denoted f^λ and is given by the hook-length formula:

$$f^\lambda = \frac{n!}{\prod_c h_c} \quad \text{where} \quad h_c = \lambda_i + \mathrm{card}\{j : \lambda_j \geq i\} - i - j + 1. \tag{1}$$

The following formulation is due to MacMahon:

$$f^\lambda = (\lambda_1 + \cdots + \lambda_d)!\det\left(\frac{1}{(\lambda_i - i + j)!}\right)_{1\leq i,j\leq d}.$$

The number of standard Young tableaux of height at most k is thus the sum

$$y_k(n) \equiv \sum_{\substack{\lambda \vdash n \\ \ell(\lambda)\leq k}} f^\lambda. \tag{2}$$

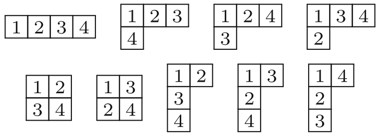
The first enumerative formulas for Young tableaux where the height is an explicit consideration appear in the 1960s, when Gordon and Houten studied k-rowed plane partitions whose nonzero parts strictly decrease along rows and columns, in addition to some related variants. In their series of Notes on Plane Partitions [20, 21], they give some formulas for the generating functions in terms of infinite products and determinants. Regev [29] first determined exact expressions for $y_2(n)$ and $y_3(n)$. The formulas are equivalent to the following. Here C_k denotes the kth *Catalan number*[1]:

$$y_2(n) = \binom{n}{\lfloor n/2 \rfloor} \qquad y_3(n) = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} C_k. \tag{3}$$

The numbers $y_3(n)$ are also known as *Motzkin numbers*. Figure 1 illustrates the nine standard Young tableaux of size four and of height at most three. Around the same time, Gessel [16] found an expression for $y_4(n)$, and Gouyou-Beauchamps [23] found the following expressions for $y_4(n)$ and $y_5(n)$:

[1] $C_n \equiv \binom{2n}{n}\frac{1}{n+1}$.

Fig. 1 All standard Young tableaux of size 4 and height at most 3



$$y_4(2n) = C_n C_n \qquad y_4(2n+1) = C_n C_{n+1}$$
$$y_5(n) = \sum_{i=0}^{\lfloor n/2 \rfloor} \frac{3!n!(2i+2)!}{(n-2i)!i!(i+1)!(i+2)!(i+3)!}. \tag{4}$$

No comparable expression for $y_6(n)$ has appeared in the literature. The presence of binomials in general, and Catalan numbers in particular, is a strong hint that these tableaux are related to well understood combinatorial classes.

1.2 The Exponential Generating Function

We study $y_k(n)$ via $Y_k(t)$, the exponential generating function for $y_k(n)$:

$$Y_k(t) \equiv \sum y_k(n) \frac{t^n}{n!}. \tag{5}$$

The formulas depends on the parity of the height. The formula for $Y_{2k}(t)$ was obtained by Gordon [19] by reducing a Pfaffian of Gordon and Houten [20], and Gessel [16] found the formula for odd heights. They are both expressed in terms of the *hyperbolic Bessel function of the first kind of order* j

$$b_j \equiv I_j(2t) = \sum_{n=0}^{\infty} \frac{t^{2n+j}}{n!(n+j)!}.$$

The formulas are:

$$Y_{2k}(t) = \det \left(b_{i-j} + b_{i+j-1} \right)_{1 \le i,j \le k} \tag{6}$$
$$Y_{2k+1}(t) = e^t \det \left(b_{i-j} - b_{i+j} \right)_{1 \le i,j \le k}. \tag{7}$$

For example, $Y_2(t) = b_0 + b_1$ and $Y_4(t) = b_0^2 + b_0 b_1 + b_0 b_3 - 2b_1 b_2 - b_2^2 - b_1^2 + b_1 b_3$. These expressions grow very fast as polynomials in b_j; however, they are amenable to some further study including asymptotic analysis of the coefficients. They also imply that $y_k(n)$ can be expressed by binomial sums, although an expression that one could compute may be too complex to be of any use.

One important consequence of these formulas is that they resolved a question that Stanley [35] had asked almost 10 years before these formulas appeared about the

nature of the function $Y_k(t)$. Specifically, he asked if $Y_k(t)$ is *D-finite*, that is, does it satisfy a differential equation with polynomial coefficients. Gouyou-Beauchamps showed that $Y_4(t)$ is D-finite from his formula, and Gessel [16] proved this result for general k since the expressions are a polynomial combination of Bessel functions (possibly times an exponential). Bessel functions are D-finite, so the result follows from closure properties.

1.3 Schur Functions

To prove these formulas, Gessel started with another important formula for standard Young tableaux. Schur functions can be described via a summation over the set of all semi-standard Young tableaux of shape λ:

$$s_\lambda = \sum_{T \in \mathrm{SSYT}(\lambda)} x^{\mathrm{content}(T)} = \sum_{T \in \mathrm{SSYT}(\lambda)} x_1^{t_1} x_2^{t_2} \dots x_k^{t_k}. \tag{8}$$

The exponents describe the content of the tableau: t_i is the number of occurrences of i in T. Thus, the coefficient of the monomial $x_1 \dots x_n$ in this expression is the number of standard Young tableaux of shape λ. We deduce the formula

$$y_k(n) = \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \le k}} [x_1 x_2 \dots x_n] s_\lambda = [x_1 x_2 \dots x_n] \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \le k}} s_\lambda.$$

This kind of coefficient extraction in symmetric functions can be framed as a homomorphism. This was done by Gessel in his PhD thesis (Theorem 3.5) and also by Jackson and Goulden [22, Lemma 4.2.5]. In the case of the homogeneous complete symmetric function h_n, it is easy to see that $[x_1 x_2 \dots x_n] h_k = \mathbb{1}_{n=k}$. One would like to apply this to the *Jacobi-Trudi* identity, which is an expression for a Schur function in terms the homogeneous complete symmetric functions:

$$s_\lambda = \det \left(h_{\lambda_i + j - i} \right)_{i,j=1}^{\ell(\lambda)}.$$

The truth is slightly more complicated. Indeed Gordon and Houten did much of the heavy lifting for this problem and had expressed the number of semi-standard tableaux as a determinant of homogenous complete symmetric functions. Gessel extracted the generating function for standard Young tableaux and derived the formulas in Eq. (6).

1.4 The Robinson–Schensted Correspondence

The following identity is incredibly evocative to combinatorialists:

$$\sum_{\lambda \vdash n} (f^{\lambda})^2 = n! \tag{9}$$

The bijective correspondence between pairs of standard Young tableaux of the same shape and permutations was described by Robinson in the 1930s and also by Schensted in the 1960s. Below we describe the Schensted algorithm which builds a pair of tableaux by parsing the permutation and incrementally building two tableaux. We refer readers to Sagan's book [31] for additional details.

An algorithmic description of the bijection begins with a permutation $\sigma \in \mathbb{S}_n$ and a pair of empty tableaux, denoted (P_0, Q_0). For i from 1 to n, we create P_i by adding a box with entry $\sigma(i)$ to P_{i-1} via a special insertion algorithm. For each i we add precisely one box. The position is noted and a box with entry i in the same location is added to Q_{i-1}. The sequence of Q tableaux records the history of the box additions. The pair (P_n, Q_n) is returned.

The row insertion takes as input a possibly incomplete standard Young tableau and adds an integer m not already in the tableau. The process acts incrementally along each row. If m is bigger than all of the elements in the row under consideration, it is placed at the end of it. If not, it finds its natural place and bumps the larger value. The bumped value is inserted in the tableau formed by the lower rows by the same process.

Example 1. Let σ be the involution $\sigma = 7\,2\,9\,6\,10\,4\,1\,8\,3\,5$. If the Schensted algorithm is applied, the penultimate step returns the following two tableaux:

$$P_9 = \begin{array}{|c|c|c|}\hline 1 & 3 & 8 \\ \hline 2 & 4 & 10 \\ \hline 6 & 9 \\ \cline{1-2} 7 \\ \cline{1-1}\end{array} \quad Q_9 = \begin{array}{|c|c|c|}\hline 1 & 3 & 5 \\ \hline 2 & 4 & 8 \\ \hline 6 & 9 \\ \cline{1-2} 7 \\ \cline{1-1}\end{array}.$$

The final step is to insert 5 into P_9. The 5 fits between the 3 and 8 in the first row, hence it bumps 8 to the next row which then bumps 10 which settles at the end of row three. The position of the new square is (3, 3), and the algorithm finishes by adding this square to Q_9, with an entry value of 10.

The end result is

$$P_{10} = Q_{10} = \begin{array}{|c|c|c|}\hline 1 & 3 & 5 \\ \hline 2 & 4 & 8 \\ \hline 6 & 9 & 10 \\ \hline 7 \\ \cline{1-1}\end{array}$$

The involution has two fixed points (2 and 8) and the resulting tableaux each have two columns of odd length (specifically, length 3).

The example illustrates some properties which are true in general. The length of the longest increasing subsequence of σ is equal to the length of the first row of the tableaux. If σ is an involution, then the number of fixed points of σ equals the number of columns of odd length in λ. That standard Young tableaux of size n are equinumerous with involutions was known to Frobenius and Schur.

Viennot [38] described a very beautiful geometric construction using *shadow lines* to give a more intuitive illustration of this slightly mechanical bijection. It should be better incorporated into the bijections we encounter in later sections. A second geometric construction is given by Fomin using growth diagrams [14]. The construction can be mined for additional information.

1.5 Plan of the Article

The first formulas found $Y_k(t)$ were recognized to resemble generating functions for walks in Weyl chambers deduced by Grabiner and Magyar [25, Section 6.2]. We examine the relevant developments made in the early 1990s [15, 39, 40, 42, 43] in Sect. 2.

Bijective proofs of some of these connections are more recent [10, 26, 33], although the work of Gouyou-Beauchamps dates back to the late 1980s. All of these authors' proofs pass through secondary objects, such as coloured Motzkin paths, or matchings. We consider these in Sect. 2.6.

One of the most common strategies for enumerating lattice walks restricted to cones with symmetry in the set of allowable steps involves a sub-series extraction from a rational function. Remarkably, many combinatorial classes with transcendental D-finite generating functions share this property. It is an open problem to answer under which conditions this might be universally true. In part, it is a useful formulation since in ideal cases we can answer questions about the order and complexity of the recurrences and also (re-) derive asymptotic formulas. We describe such extractions, and the resulting tableaux generating functions in Sect. 3.

From the Robinson–Schensted correspondence, we see that the combinatorial cousin to the standard Young tableau of bounded height is the permutation with bounded longest increasing subsequence. Many of the techniques described here can also enumerate these classes. For example, Gessel determined expressions for the generating functions, and they also have lattice walk interpretations. We summarize results in Sect. 4.

We conclude with some natural generalizations and open problems.

2 Lattice Walk Models

There are no fewer than five different lattice walk classes that are equinumerous with standard Young tableaux of bounded height. Most of them are defined using d-dimensional Weyl chambers[2] of type C, denoted by $W_C(d)$

$$W_C(d) \equiv \{(x_1, x_2, \ldots, x_d) : x_1 \geq x_2 \geq \cdots \geq x_d \geq 0\}.$$

Let $\{e_1, \ldots, e_d\}$ denote the standard basis of $\mathbb{R}^d$. A lattice walk model is defined by a set of allowable steps and a region which confines each walk. A walk is a sequence of steps. For example, the set

$$\mathcal{S} = \{\pm e_i : 1 \leq i \leq d\}$$

defines the d dimensional *simple* step set. This is considered in several different bounding cones.

Gessel and Zeilberger [17] considered general walks in Weyl chambers and demonstrated an enumeration strategy for models where the stepset possesses a certain kind of symmetry, and avoid jumping over boundaries. Such walks have come be called walks *reflectable*. The generating function for reflectable walk models with specified starting and endpoints (excursions) is written as a coefficient extraction from a signed sum of unrestricted walks. This idea, which appears frequently in lattice walk enumeration, is a version of the reflection principle. The expressions were made explicit for $W_C(d)$ by Grabiner and Magyar for the simple step set, which happens to be reflectable.

Theorem 1 (Grabiner and Magyar [25]). *For fixed $\lambda, \mu \in W_C(d)$, the exponential generating function $O_{\lambda,\mu}(t)$ of the simple walks in $W_C(d)$ from λ to μ, counted by their lengths, satisfies*

$$O_{\lambda,\mu}(t) = \det\left(b_{\mu_i-\lambda_j} - b_{\mu_i+\lambda_j}\right)_{1\leq i,j\leq d}.$$

This is immediately reminiscent of the generating function formulas in Eq. (6). Several authors have made the connection, in particular, Gessel, Weinstein and Wilf [15], Zeilberger [43], Xin [42], Eu et al. [32, 33], Burrill et al. [10] and Courtiel et al. [12]. In almost every case, there is a natural parameter which is equidistributed with number of odd columns in the tableaux. We describe these classes next.

[2] For convenience we define the chambers using non-strict inequalities, our bijective statements can equivalently be given under strict inequalities, upon applying the coordinate shift $\widetilde{x}_i = x_i + d + 1 - i$.

2.1 Ballot Walks

We can build an integer sequence from the entries of a standard Young tableaux. We define the associated lattice word $w = (w_1, w_2, \dots, w_n)$ by setting $w_i = j$ if the entry i is in the jth row. This word has the property that for any prefix the number of occurrences of ℓ is greater than or equal to the number of occurrences of $\ell + 1$ since the columns of the associated tableau are strictly decreasing. These are also known as generalized Ballot sequences.

We associate a step naturally to each letter:

$$w_1 \to e_1, \qquad w_i \to e_i - e_{i-1} \quad \text{for } (2 \le i \le k-1), \qquad w_k \to -e_{k-1}.$$

Restricting walks to the first orthant is equivalent to the ballot condition. That is, the ballot word of a tableaux of height at most k gives a natural encoding as a lattice walk in the cone $\mathbb{R}_{\geq 0}^{k-1}$ with steps from the following stepset:

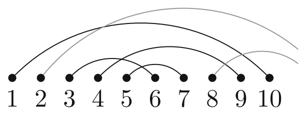
$$\mathcal{B} \equiv \{e_1, -e_{k-1}\} \cup \{e_i - e_{i-1} : 2 \le i \le k-1\}.$$

Figure 2 has an example.

$$\iota = \begin{array}{|c|c|c|} \hline 1 & 3 & 5 \\ \hline 2 & 4 & 8 \\ \hline 6 & 9 & 10 \\ \hline 7 & \multicolumn{2}{c}{} \\ \cline{1-1} \end{array}$$

$$\sigma = \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 7 & 2 & 9 & 6 & 10 & 4 & 1 & 8 & 3 & 5 \end{array}$$

$$\omega = (b_1, b_2, b_1, b_2, b_1, b_3, b_4, b_2, b_3, b_3)$$

$\mu =$ 

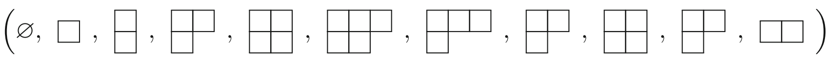
$\theta = ($  $)$

Fig. 2 Representatives of classes of objects in bijection with standard Young tableaux of bounded height, in particular their image of ι, a standard Young tableau of height 3 with two odd columns; σ is an involution with maximal increasing sequence of length 3; ω is a ballot walk; μ is an open arc diagram for a partial matching; θ is an oscillating tableau ending on a row shape

2.2 Lazy Walks

To define this step set, we set $\overline{0}$ to be the zero step (whence the label "lazy"). Zeilberger [43] noted (although, he attributes the proof to Xin without citation) that the number of excursions in $W_C(k)$ starting and ending at the origin using the stepset $\mathcal{L}$, given by

$$\mathcal{L} \equiv \{e_i, -e_i : 1 \leq i \leq k\} \cup \{\overline{0}\}$$

is $y_{2k+1}(n)$. A small computation suggests that the distribution of the zero steps matches the distribution of odd columns in the Young tableaux. He remarks that it would be interesting to find a (bijective) proof of this result.

2.3 Generalized Motzkin Paths

In Zeilberger's lazy walks, the $k = 1$ case encodes the classic Motzkin walks, consistent with the longstanding observation that $y_3(n)$ is the number of Motzkin words. One could view the higher dimension lazy walks as a generalization of Motzkin words, and this notion was first formalized by Eu [32], and subsequently by Eu, Fu, Hou, and Hsu [33]. They add a counter component and describe an explicit bijection between the Motzkin paths of length n and the standard Young tableaux of size n with at most three rows.

To prove this, they considered the lazy lattice walks and then algorithmically mapped the steps in. The odd and even cases reconcile as follows.

Theorem 2 (Eu, Fu, Hou, and Hsu 2013 [33, Theorem 1.1])**.** *Consider the lattice model defined by the step set*

$$\mathcal{M} \equiv \{e_1\} \cup \{e_1 + e_2, e_1 - e_2\} \cup \{e_1 - e_i + e_{i+1}, e_1 + e_i - e_{i+1} : 2 \leq i \leq k\}$$

confined to $\mathbb{Z}_{\geq 0}^{k+1}$*. The number of walks of length n from the origin to the point* $(n, 0, ..., 0)$ *staying within the nonnegative octant equals the number of n-cell SYTs with at most* $2k + 1$ *rows.*

If, additionally, the e_1 steps are confined to the hyperplane spanned by the vectors $\{e_1, \ldots, e_k\}$*, then the number of paths equals the number of n-cell standard Young tableaux with at most* $2k$ *rows.*

The number of e_1 steps is equidistributed with the odd column statistic of standard Young tableaux.

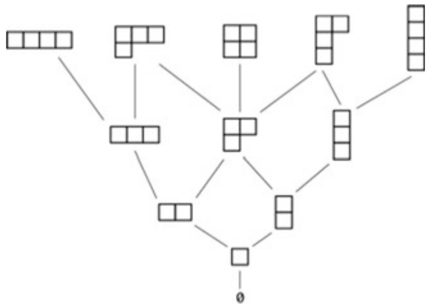


Fig. 3 The first few levels of Young's lattice of Ferrers diagrams

2.4 Oscillating Tableaux and Arc Diagrams

The set of Ferrers diagrams ordered by diagram inclusion[3] is called Young's lattice. Figure 3 depicts the first few levels of its Hasse diagram. We consider a sequence of Ferrers diagrams as a walk on this lattice. We consider three variants defined by restrictions on moving up or down in the lattice (or not at all). The *length* of a sequence is the number of elements, minus one. (It is the number of steps in the corresponding walk.)

An *oscillating tableau* is simply a sequence of Ferrers diagrams such that at every stage a box is either added or deleted. They were popularized by their use in interpretations for representations of the symplectic group [36]. If no diagram in the sequence is of height $k + 1$, we say that the tableau has its *height bounded by k*. We recall two related families, namely *vacillating tableaux*, and *hesitating tableaux*. The vacillating tableaux are *even* length sequences of Ferrers diagrams, written $(\lambda^{(0)}, \ldots, \lambda^{(2n)})$ where consecutive elements in the sequence are either the same or differ by one square, under the restriction that $\lambda^{(2i)} \geq \lambda^{(2i+1)}$ and $\lambda^{(2i+1)} \leq \lambda^{(2i+2)}$. The hesitating tableaux are even length sequences of Ferrers diagrams, written $(\lambda^{(0)}, \ldots, \lambda^{(2n)})$ where consecutive differences of elements in the sequence are either the same or differ by one square, under the following restrictions:

- if $\lambda^{(2i)} = \lambda^{(2i+1)}$, then $\lambda^{(2i+1)} < \lambda^{(2i+2)}$ (do nothing; add a box)
- if $\lambda^{(2i)} > \lambda^{(2i+1)}$, then $\lambda^{(2i+1)} = \lambda^{(2i+2)}$ (remove a box; do nothing)
- if $\lambda^{(2i)} < \lambda^{(2i+1)}$, then $\lambda^{(2i+1)} > \lambda^{(2i+2)}$ (add a box; remove a box).

Figure 4 shows examples from each of these classes.

Chen, Deng, Du, Stanley, and Yan [11] described nontrivial bijections between sequences of Ferrers diagrams and several combinatorial families encoded in arc diagrams. *Arc diagrams* are labelled graphs under some degree and embedding restraints. They can be used to represent a variety of combinatorial classes, such as matchings and partitions. Figure 5 illustrates how to encode a set partition as an arc diagram.

[3]Recall $\lambda \leq \mu$ means that $\lambda_i \leq \mu_i$ for all i.

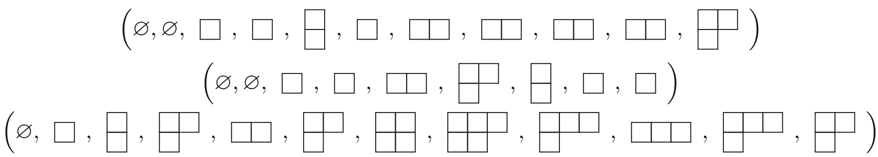


Fig. 4 *From top to bottom:* a vacillating tableau of length 10; a hesitating tableau of length 8; an oscillating tableau of length 11. In each case, the height is bounded by 2. From [10]

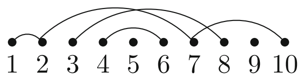


Fig. 5 An arc diagram representation of the set partition $\pi = \{1, 2, 7, 10\}, \{3, 8\}, \{4, 6\}, \{5\}, \{9\}$. It is both 3-noncrossing and 3-nonnesting. However, the subdiagram induced by $\{2, 3, 7, 8, 10\}$ is an enhanced 3-crossing. Similarly, the subdiagram induced by $\{3, 4, 5, 6, 8\}$ is an enhanced 3-nesting

They generalize the Schensted algorithm, in some sense, by describing how to parse arc diagrams, with each step defining a tableau insertion or deletion. The result is bijection ϕ from partitions to vacillating tableau. It is robust and can be adapted to other arc diagram classes. A key feature is sub-pattern avoidance properties in the arc diagrams are mapped to height restrictions on the tableaux. This echoes a key feature of the Schensted algorithm.

An arc diagram is noncrossing if no two arcs intersect. Noncrossing set partitions are counted by Catalan numbers. In addition to appearing in combinatorics, these diagrams arise in algebra, physics, and free probability. The notion of a crossing is generalized to a k-*crossing*, which denotes a set of k arcs that each mutually cross. Similarly, a k-*nesting* refers to k arcs which mutually nest into a rainbow figure. More formally, let us consider a set of k distinct arcs: $(i_1, j_1), \ldots, (i_k, j_k)$. They form a k-crossing if $i_1 < i_2 < \cdots < i_k < j_1 < j_2 < \cdots < j_k$, and a k-nesting if $i_1 < i_2 < \cdots < i_k < j_k < \cdots < j_2 < j_1$. A slightly relaxed notion is sometimes appropriate. Enhanced k-nestings and k-crossings permit the middle inequality to be an equality. We are interested in the classes of diagrams which *avoid* such sub-diagrams and consider k-noncrossing diagrams (they contain no k-crossing) and k-nonnesting diagrams (which contain no k-nesting). Figure 5 gives examples.

Theorem 3.2 of Chen et al. [11] proves that given a partition π of $\{1, \ldots n\}$, and the vacillating tableau $\phi(\pi) = (\emptyset = \lambda^{(0)}, \lambda^{(1)}, \ldots, \lambda^{(2n)} = \emptyset)$, the size of the largest crossing of ϕ is the largest height of any λ_i, and the size of the largest nesting is the maximum number of columns.

Thus, by a simple transposition one deduces the corollary comprised of the surprisingly nontrivial fact that k-noncrossing set partitions of $\{1, 2, \ldots, n\}$ are in bijection with k-nonnesting set partitions of $\{1, 2, \ldots, n\}$. We leave it to the reader to see why a simple swap of sub-diagrams is not a bijection.

The bijection ϕ supports many generalizations. A diagram is said to be *open* if, in addition to closed edges, there are also half edges. The notion of crossing extends naturally, given a fixed convention of how the open edges lie. The diagram μ in Fig. 2

demonstrates an incomplete, or open, matching. Remark there are not three edges that mutually cross, and hence, this diagram is also 3-noncrossing. An open diagram corresponds to a sequence that does not necessarily end at an empty diagram—rather it ends in a row shape, i.e. a partition with a single part: (m). The bijection ϕ is well defined in this case, although Chen et al. did not discuss how to interpret this, although Burrill, Melczer and Mishna [9]. Xin [42] consider walks with any endpoint and found generating functions for palindromic sequences.

In a sequence of Ferrers diagrams of height at most k, each diagram can be coded by a k-dimensional vector where each successive entry is weakly decreasing. Thus, a sequence is an encoding of a walk in $W_C(d)$. Thus, we can view the bijection ϕ as a nontrivial map between arc diagrams and walks. To end in a row shape is the same as ending on the first axis.

This brings us back to our topic at hand, standard Young tableaux. Which arc diagrams are associated with a standard Young tableau of bounded height? The most plain interpretation is to view a standard Young tableau as a sequence of Ferrers diagrams—it starts at the empty diagram and a single cell is added at each step, with the position indicated by the entry in the tableaux. We could consider the arc diagram image of this sequence, but we can do better.

Recall the Robinson–Schensted correspondence is a bijection between pairs of standard Young tableaux and permutations. By restricting the map to pairs of identical tableaux, it becomes a bijection between standard Young tableaux and involutions. Involutions have a very natural arc diagram representation! In this correspondence, a tableau with a fixed number of odd columns is mapped to an involution with precisely that many fixed points. An involution is a partial matching, and they are in bijection with oscillating tableaux.

These pieces lead to the following result, a map between standard Young tableaux and a class of walks that end on a boundary.

Theorem 3 (Krattenthaler [26, Theorem 4]; Burrill, Courtiel, Fusy, Melczer and Mishna [10, Theorem 1]; Okada [28, Theorem 1.2])**.** *For $n, k \geq 1$, there is an explicit bijection between the standard Young tableaux of size n with height at most k and with m odd columns, and the simple walks of length n staying in $W_C(k)$, starting from the origin and ending at the point $m\,e_1$.*

We have outlined one possible argument, but in fact three very different proofs have been discovered. Above is the argument of Burrill, Courtiel, Fusy, Melczer and Mishna [10]. Krattenthaler [26] determines a different bijection using growth diagrams as an intermediary object. It is direct and gives several generalizations. It is self-contained in that it does not exploit the τ map. Okada [28] uses a representation theoretic argument starting from Pieri rules. He also has generalizations to semi-standard tableaux. Gouyou-Beauchamps proved the $k = 4$ case.

One consequence of a lattice walk bijection is a generating function expression. Applying Grabiner and Magyar's formula gives the following

$$Y_k(t) = \sum_{u=0}^{k-1} (-1)^u \sum_{\ell=u}^{2k-1-2u} (b_\ell) \det(b_{i-j} - b_{kd-i-j})_{0\leq i\leq k-1, i\neq u, 1\leq j\leq k-1}.$$

Remarkably, the infinite sum which arises from direct application of Grabiner and Magyar's formula telescopes into a finite sum. We also use the identity $b_{-k} = b_k$.

2.5 Excursions in the Weyl Chamber of Type D

Classically, there is a simple bijection between walks with steps from $\{(1, 1), (1, -1)\}$ which, in the first instance start and end on the axis, with no further restriction (let us call them bridges), and in the second instance start at the origin, and never go below the axis and can end at any height (Dyck prefixes). The first class restricts free walks by restricting the possible endpoints, and the second restricts free walks by restricting the size of the region. We can view this as a trade-off of restrictions. The simplest proof of this result passes through a third object, marked Dyck paths, which are objects in the intersection of both classes, but have additional markings on some of the down steps which touch the axis. Consider the following two maps, which both define bijections. A Dyck path with marked down steps is mapped to a Dyck prefix, by flipping all of the marked down steps into upsteps. The flipped marked steps become the last step at that height in the walk, and the Dyck prefix ends at a height equal to the number of marked steps. On the other hand, we could consider the entire segment of the walk ending at a marked down step, starting at the nearest previous up step which touched the axis. We can flip this entire segment across the axis. We do this for every marked step to build a bridge. The marked intermediary facilitates a straightforward proof, but how to mark in higher dimensions?

The Weyl chamber of type D is the following region:

$$W_D(k) \equiv \{(x_1, x_2, \ldots, x_k) \ : \ x_1 \geq x_2 \geq \cdots \geq x_{k-1} \geq |x_k|\}\,.$$

An *axis walk* is any walk starting at the origin and ending on the x_1-axis.

Theorem 4 (Courtiel, Fusy, Lepoutre and Mishna [12, Theorem 20]). *For $k \geq 1$ and $n \geq 0$, there is an explicit bijection between simple axis walks of length n staying in $W_C(k)$ and simple excursions of length n staying in $W_D(k)$, starting from $(\frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{2})$, and ending at $(\frac{1}{2}, \ldots, \frac{1}{2}, \frac{(-1)^n}{2})$. The ending x_1-coordinate of a walk from $W_C(k)$ corresponds to the number of steps that change the sign of x_k in its bijective image.*

The proof of this result is a generalization of the marked Dyck path example. The analog to the Dyck prefix is the axis walk in the orthant space $W_C(k)$, and the analog to the bridge is the excursion in the larger space $W_D(k)$. The intermediary class, marked excursions in $W_C(k)$, are defined by a less straightforward process. For the first bijection, Courtiel et al. map the axis walks to open arc diagrams. The open arcs are removed, but their position is marked. The inverse bijection is applied and a marked excursion results. The second bijection is slightly more complicated, and the reader is referred to the article for details.

A corollary of this result is another lattice model in bijection with standard Young tableaux of bounded height.

Corollary 5 (Courtiel, Fusy, Lepoutre, Mishna [12, Corollary 21]). *For $n, k \geq 1$, there is an explicit bijection between the standard Young tableaux of size n with height at most k, and the simple walks of length n staying in $W_D(k)$, starting from $(\frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{2})$, and ending at $(\frac{1}{2}, \ldots, \frac{1}{2}, \frac{(-1)^n}{2})$. The number of odd columns corresponds to the number of steps that change the sign of x_k.*

It seems very likely that an explicit bijection could be found between these and the lazy walks. This would answer the question of Zeilberger and likely reveal a model for the odd cases.

There is also potential for another generating function formula by applying the results of Grabiner and Magyar.

2.6 A Collection of Bijections

We summarize some combinatorial classes in bijection with standard Young tableaux of bounded height. Figure 2 illustrates some of these classes.

Theorem 6. *The set of standard Young tableaux of size n with height bounded by k and m odd columns is in bijection with each of the following sets:*

(1) The set of involutions of size n with m fixed points and no decreasing subsequence of length $k+1$;
(2) The set of oscillating tableaux of size n with height bounded by k, which start at the empty partition and end in a row shape $\lambda = (m)$;
(3) (if k is even) The set of open matching diagrams of length n, with m open arcs and with no $(k/2+1)$-crossing;

Gil, McNamara, Tirrell, and Weiner [18, Theorem 1.1] proved the equivalence between Class (3) in the theorem, and a set of Dyck paths where each group of ascent steps are decorated by a connected matching of a related size. A main feature of their construction is the following identity. Suppose $P_k(t)$ denotes the generating function for the number of k-noncrossing perfect matchings on $[2n]$, then

$$Y_{2k-1}(t) = \frac{1 + P_k(t^2(1-t)^{-2})}{1-t}.$$

3 Generating Function Expressions

The generating functions of lattice walk in cones with symmetries in the stepset can often be expressed as a sub-series extraction of a Laurent expansion of a rational

function. This generalizes the result on reflectable walks of Gessel and Zeilberger and is at the heart of the orbit sum method of Bousquet-Mélou and Mishna [6] which handles a wide class of lattice walks restricted to the quarter plane. In their work, they start with a natural combinatorial recurrence on lattice walks, which translates to a functional equation satisfied by the generating function. For some models, it is sufficient to take a weighted sum (the namesake orbit sum) of the equations and isolate the target generating function with a sub-series extraction.

In this context, the extraction operates on iterated Laurent series. For a function $F(x_1, \ldots, x_d; t) \in \mathbb{C}(x_1, \ldots, x_d)(t)$ which is analytic at the origin, we denote by $[x_1^{k_1} x_2^{k_2} \ldots x_d^{k_d} t^n] F(x_1, \ldots, x_d; t)$ to be the coefficient of the term $x_1^{k_1} x_2^{k_2} \ldots x_d^{k_d} t^n$ in the Laurent expansion. In this context, we view the objects to be series in t variable. We recall the terminology that several authors use for the special case of the constant term with respect to a set of variables: $\mathrm{CT}_{x_1 \ldots x_d} F(x_1, \ldots, x_d; t)$ is the coefficient of the term $x_1^0 \ldots x_d^0$ in the Laurent series expansion of F at the origin. It is a series in the remaining variables. It can be obtained by incrementally determining the constant term with respect to a single variable. Here we shall take the order of the variables as they are listed, for example. Such a coefficient extraction can be written as a Cauchy integral.

A related operator is the *diagonal*. The (central) diagonal $\Delta F(x_0, x_1, \ldots, x_d)$ of a formal power series is the univariate sub-series defined

$$\Delta F(x_0, \ldots, x_d) = \Delta \sum_{i_0, \ldots, i_d} f(i_0, \ldots, i_d) x_0^{i_0} \ldots z_d^{i_d} \equiv \sum_n f(n, \ldots, n)\, t^n.$$

Lipshitz [27] proved that the diagonal of a D-finite function[4] is also D-finite. Since rational functions are D-finite, diagonals of rational functions are also D-finite. The expressions we consider here are all diagonals of rational functions and hence are all D-finite by construction. This provides an alternate proof of the D-finiteness of $Y_k(t)$.

Bousquet-Mélou [7] used an orbit sum to directly derive generating function expressions for standard Young tableaux of bounded height. Her generating functions also mark the number of odd columns, and other parameters are also readily accessible. In that same paper, she determines a nice functional equation proof of the hook length formula. The starting point is a simple recursive construction for standard Young tableaux: a tableau of size $n + 1$ is obtained from a tableau of size n by adding a cell to the jth row, unless the jth row is already the same length as the $j - 1$ row. This translates into a very straightforward functional equation for the generating function

$$F(u) \equiv \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq k}} f^\lambda u_1^{\lambda_1} u_2^{\lambda_2} \cdots .$$

[4] A (multivariate) function is D-finite if the set of all its partial derivatives spans a vector space of finite dimension.

The equation is defined in term of $F_j(u)$, the generating function for those standard Young tableaux such that $\lambda_{j-1} = \lambda_j$:

$$F(u) = 1 + u_1 F(u) + \sum_{j=2}^{k} u_j (F(u) - F_j(u)).$$

She applies a kernel method argument to this functional equation to recover MacMahon's formula for f^λ.

She has generating function expressions in her Propositions 9 and 10. It is her Proposition 11 that accounts for the additional parameter of interest. We note that she works with the ordinary generating function,

$$\widetilde{Y}_k(u; t) \equiv \sum_n \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \le k}} f^\lambda\, u^{\#\text{odd columns in } \lambda}\, t^n.$$

Theorem 7 (Bousquet-Mélou [7]). *For $k = 2\ell$, the ordinary generating function of standard Young tableaux with height bounded by k where t marks the length and x_1 marks the number of odd columns is*

$$\widetilde{Y}_k(x_1; t) = \mathrm{CT}_{x_2,\ldots,x_k} \frac{-\det(x_j^{-i} - x_j^i)_{1 \le i,j \le \ell}}{1 - t(x_1 + x_1^{-1} + \cdots + x_\ell + x_\ell^{-1})} \frac{1}{x_2 x_3^2 \ldots x_\ell^{\ell-1}}. \tag{10}$$

The orbit sum strategy can be used to determine an expression for the ordinary generating function for simple walks in $\mathcal{W}_C(k)$ ending on the axis. This gives the following result.

Theorem 8 (Burrill, Melczer and Mishna [9]). *The ordinary generating function, $\widetilde{Y}_k(t)$ for Young tableaux of height bounded by k satisfies the formula formula*

$$t^{2k-1}\widetilde{Y}_{2k}(t) = -\Delta\left[\frac{z_0^{2k-1}(z_3 z_4^2 \cdots z_k^{k-2})(z_1+1)\prod_{1\le j<i\le k}(z_i - z_j)(z_i z_j - 1)\cdot\prod_{2\le i\le k}(z_i^2 - 1)}{1 - z_0(z_1\cdots z_k)(z_1 + z_1^{-1} + \cdots z_k + z_k^{-1})}\right].$$

These formulas are not necessarily easier to compute or interpret than the others, but they do have a few implications. They do provide a proof that the generating function is D-finite. Furthermore, we see a second proof that $y_k(n)$ can be expressed as binomial sums, and potentially a different route to obtain such expressions, using the work of Bostan, Lairez and Salvy [5]. They may be suitable for analysis by the methods of Pemantle and Wilson [41], which determine asymptotic formulas for coefficients of functions which are expressed as diagonals of rational functions.

Since it is the generating function for Motzkin numbers, $\widetilde{Y}_3(t)$ is algebraic. Because the asymptotics of the coefficients are incompatible with algebraicity, $\widetilde{Y}_4(t)$ is not. For which k is the ordinary generating function $\widetilde{Y}_k(t)$ algebraic?

3.1 Differential Equations

Once we know that a generating function is D-finite, it is natural to ask about differential equations that it satisfies. Bergeron, Favreau and Krob [3] generated many conjectures about the order of the differential equation satisfied by $Y_k(t)$ from computer experiments, and some analysis of its expression as a determinant of a matrix of modified Bessel functions.

Proposition 1 in [2] states the dimension of the vector space over the field $\mathbb{C}(t)$ of rational functions in t spanned by $Y_k(t)$ and all its derivative is bounded by $\lfloor \frac{k}{2} \rfloor$.

Conjecture 9 (Bergeron, Favreau, Krob [3]). *For each k, there are polynomials $p_m(t)$ of degree at most $\lfloor \frac{k}{2} \rfloor$ such that $Y_k(t)$ is a solution of a linear differential equation order at most $\lfloor \frac{k}{2} \rfloor + 1$ with coefficients $p_m(t)$.*

They have verified this conjecture for $k \leq 11$.

We can use recent advances in symbolic computation and the diagonal expression in Theorem 8 to bound the order and degree of the differential equations satisfied by the ordinary generating function $\widetilde{Y}_k(t)$. In particular, the following theorem of Bostan, Lairez, and Salvy [4] can be explicitly applied.

Theorem 10 (Bostan, Lairez, and Salvy [4]). *Let $R(z_1, \ldots, z_d, t) = A(\mathbf{z}, t)/B(\mathbf{z}, t) \in \mathbb{Q}(t)(z_1, \ldots, z_d)$, be a rational function with multidegree bounds*

$$n_{\mathbf{z}} := \max(\deg_{\mathbf{z}} B, \deg_{\mathbf{z}} A + d + 1) \qquad n_t := \max(\deg_t A, \deg_t B).$$

Then there exists an annihilating differential equation $\mathcal{L}$ for the integral

$$P(t) := \oint_{\gamma} R(\mathbf{z}, t) d\mathbf{z},$$

where γ is any n-cycle in $\mathbb{C}^n$ on which R is continuous when t ranges over some connected open set $U \subset \mathbb{C}$ (note that $\mathcal{L}$ is independent of γ). Furthermore the order of $\mathcal{L}$ is at most $n_{\mathbf{z}}^d$ and the degree of $\mathcal{L}$ is at most $\left(\frac{5}{8} n_{\mathbf{z}}^{3d} + n_{\mathbf{z}}^d\right) e^d n_t$.

We write a diagonal as an integral:

$$\Delta F(\mathbf{z}, t) = \left(\frac{1}{2\pi i}\right)^d \int_{\gamma} \frac{F(z_1, z_2/z_1, z_3/z_2, \ldots, z_d/z_{d-1}, t/z_d)}{z_1 z_2 \cdots z_d} d\mathbf{z} \qquad (11)$$

by the multivariate Cauchy residue theorem, for an appropriate n-cycle γ around the origin.

In the case of $\widetilde{Y}_k(t)$, we compute an upper bound of $2d^2 - 3d + 1$ on the total degree of the denominator in the $\mathbf{z}$ variables and a bound of $2d^2 - 4d - 2$ on the degree of the numerator and implies the following result, computed by Melczer.

Proposition 11. *The generating function for the number of standard Young tableaux of height at most k satisfies a linear differential equation of order at most $(2k^2 - 3k + 1)^k$ and degree at most*

$$\left(\frac{5}{8}(2d^2 - 3d + 1)^{3d} + (2d^2 - 3d + 1)^d\right) e^d.$$

3.2 Asymptotics

Regev [29] determined asymptotic expansions for the number of Young tableaux of bounded height. He explicitly deduced asymptotics for the $k = 3$ case as a particular extraction:

$$(n+1)y_3(n) = [x^{-1}](x + 1/x + 1)^{n+1} \sim \sqrt{\frac{3}{8}\pi} \cdot \frac{1}{\sqrt{n}} 3^n.$$

By a slightly more general argument, he showed

$$y_{2k}(n) \sim_{n\to\infty} (2/\pi)^{k/2}(2k)^n(k/n)^{k(k-1/2)} \prod_{i=0}^{k-1}(2i)!. \tag{12}$$

The asymptotics of lattice walks in cones have been well studied. The tour de force of Denisov and Wachtel [13] describes a collection of very comprehensive results. The formulas given in [13, Theorem 6] should be applied here, for example to the lazy walks and to the excursions in $W_D(k)$.

Grabiner [24] used his formulas for walks in Weyl chambers to find the asymptotics of the probability that a randomly chosen standard Young tableau of size n with at most t rows contains a given subtableau. This is equivalent to counting walks that have visited a particular point—there might be similar results to be extracted by considering other lattice models.

The asymptotic formula that has been described so far are valid for fixed k, as n tends to infinity. It is open to develop formulas when k is a function of n.

4 Restricting Increasing Subsequences in Permutations

Motivated by algebraic interpretations, Regev [29] considered the quantity

$$y_k^{(\beta)}(n) \equiv \sum_{\lambda\in\mathcal{P}_k} (f^\lambda)^\beta$$

What can be said of the generating functions

$$\sum y_k^{(\beta)}(n)t^n \quad \sum y_k^{\beta}(n)t^n u^{\beta}?$$

When $\beta = 2$, this counts permutations with restricted longest increasing subsequence, and it is very well studied, with many relevant connections to lattice walks [39, 40]. Define $u_k(n)$ to be the number of pairs of Young tableaux of the same shape with at most k rows. By the Robinson–Schensted correspondence, this is the number of permutations in $\mathfrak{S}_n$ with no $(k+1)$-increasing subsequence. For every $k \geq 1$, Gessel [16] proved the formula

$$\sum_{n\geq 0} \frac{u_k(n)}{n!^2 t^{2n}} = \det\left(b_{i-j}\right)_{1\leq i,j\leq k}, \tag{13}$$

where the b_j are the Bessel function evaluations defined earlier.

A combinatorial proof of this expression has been given by Gessel et al. [15] via simple walks ending at so-called Toeplitz points, and Xin [42] did a combinatorial derivation based on arc diagrams, and yet another constant term extraction. The permutations satisfy a nice combinatorial recursion, and, in a manner similar to her solution for involutions, Bousquet-Mélou [7] determines functional equations that can be resolved using a kernel method approach.

She determined [7, Proposition 13] the following expression for the ordinary generating function of permutations avoiding the pattern $1\,2\ldots m\,(m+1)$:

$$U_k(t) \equiv \sum_{n\geq 0} u_k(n)t^{2n} = \mathrm{CT}_{x_1,\ldots,x_m} \frac{\det((x_j - x_{j-1})^{i-j})_{1\leq i,j\leq m}}{1 - t\left(\sum \frac{1}{x_j - x_{j-1}}\right)} \cdot \sum_{i=0}^{m}\prod_{j=1}^{i} \frac{x_j}{1-x_j}. \tag{14}$$

Bergeron and Gascon found the differential equations satisfied by the exponential generating functions for $k < 11$. It remains open to answer if $(y_k^{(\beta)}(n))$ is the counting sequence for any easily characterizable combinatorial family, perhaps as a restricted or decorated family of permutations. The argument of Gessel [16] applies to prove that the sequence is P-recursive for positive integer β, and probably also a diagonal of a rational function. Can we usefully bound the annihilating differential operators or determine asymptotic formulas for arbitrary β?

Wilf [40] deduced $U_{2k}(t) = Y_{2k}(t)Y_{2k}(-t)$. From this, it follows

$$\binom{2n}{n} u_{2k}(n) = \sum_r \binom{2n}{r}(-1)^r y_{2k}(r)y_{2k}(2n-r).$$

Is there a lattice path interpretation of this identity? Are there identities for other β values?

5 Other Directions

5.1 Using Kronecker Coefficients

The Kronecker product of symmetric functions gives an important connection to representation theory. In particular, the Kronecker product of two symmetric functions in the Schur function basis determines the multiplicities of irreducible characters in this tensor product.

The following Schur function identity for the Kronecker product of two Schur functions (denoted by $*$) was shown by Brown, van Willigenburg and Zabrocki [8]:

$$s_{(n,n-1)} * s_{(n,n-1)} = \sum_{\substack{\lambda \vdash 2n-1 \\ \ell(\lambda) \leq 4}} s_\lambda.$$

Tewari [37] manipulates this formula to deduce a closed form for the number of Young tableaux with height exactly 5, under the additional constraint of $\lambda_5 = 1$. His Theorem 7.4 is a simple sum of Motzkin numbers (recall $M_n = y_3(n)$) and pairs of Catalan numbers (recall also the formula of y_4):

$$\sum_{\substack{\lambda \vdash 2n \\ \ell(\lambda)=5 \\ \lambda_5=1}} f^\lambda = \left(\frac{n(n+2)}{2n+1}\right) C_n C_{n+1} - C_{n+1}^2 + M_{2n},$$

$$\sum_{\substack{\lambda \vdash 2n-1 \\ \ell(\lambda)=5 \\ \lambda_5=1}} f^\lambda = \left(\frac{(n+1)}{2}\right) C_n^2 - C_n C_{n+1} + M_{2n-1}.$$

It could be straightforward to find a combinatorial interpretation of his formula, which then could be generalized to higher dimensions. It also suggests that perhaps some of the larger summations could be simplified. This could be key to finding a useful expressions for $y_6(n)$.

5.2 Other Classes in Bijection

As Gouyou-Beauchamps noted in [23], the numbers that appear in $y_4(n)$ also appear in the enumeration of planar maps and alternating Baxter permutations. Baxter permutations are a class of pattern avoiding permutations that are very combinatorially rich. They also have a bijection to lattice path models, using the machinery of arc diagrams as an intermediary class. It might be possible to directly connect these classes.

5.3 Shadow Diagrams

Some of the lattice walk bijections use tableau insertion and deletion in more than one stage. Perhaps there exist more economical, or direct bijections. The shadow diagrams of Viennot may play an important role in such a simplification.

5.4 Random Tableaux

There have been several recent works on generating random walks. One application is to convert this to a generator for random tableaux. Which of the above bijections has smallest complexity?

Grabiner [24] was able to use lattice walk results to determine distributions of subtableaux in Young tableaux. It seems that there should be more results along this vein with each of the lattice walk representations.

5.5 Semi-standard Young Tableaux

Okada [28] proved some results connecting counting sequences of generalized oscillating tableaux and semi-standard tableaux using techniques from representation theory. In particular, his Theorem 5.3 is a list of results on equinumerous classes of tableaux of bounded height that are well suited for more bijective explanations.

Krattenthaler's results [26] on semi-standard tableaux replace single steps in oscillating tableaux with jumps by horizontal strips. Perhaps they can connect here using lattice walks with longer steps or diagonal steps. In any case, these two works should be connected more explicitly.

Acknowledgements The author is grateful to MSRI for travel support to participate in the 2017 AWM session. This expository work was inspired by that meeting. I am grateful for the patience and wisdom of the anonymous referees. The author's research is also partially funded by NSERC Discovery Grant RGPIN-04157.

References

1. R. Adin, Y. Roichman, Standard Young tableaux, *Handbook of Enumerative Combinatorics*, Discrete Mathematics and its Applications (CRC Press, Boca Raton, 2015), pp. 895–974
2. F. Bergeron, F. Gascon, Counting Young tableaux of bounded height. J. Integer Seq. **3**(2), 3 (2000)
3. F. Bergeron, L. Favreau, D. Krob, Conjectures on the enumeration of tableaux of bounded height. Discret. Math. **139**(1–3), 463–468 (1995)

4. A. Bostan, P. Lairez, B. Salvy, Creative telescoping for rational functions using the Griffiths-Dwork method, in *ISSAC 2013—Proceedings of the 38th International Symposium on Symbolic and Algebraic Computation*, (ACM, New York, 2013), pp. 93–100
5. A. Bostan, P. Lairez, B. Salvy, Multiple binomial sums. J. Symb. Comput. **80**(part 2), 351–386 (2017)
6. M. Bousquet-Mélou, M. Mishna, Walks with small steps in the quarter plane, in *Algorithmic Probability and Combinatorics*, vol. 520, Contemporary Mathematics (American Mathematical Society, Providence, 2010), pp. 1–39
7. M. Bousquet-Mélou, Counting permutations with no long monotone subsequence via generating trees and the kernel method. J. Algebr. Comb. **33**(4), 571–608 (2011)
8. A.A.H. Brown, S. van Willigenburg, M. Zabrocki, Expressions for Catalan Kronecker products. Pac. J. Math. **248**(1), 31–48 (2010)
9. S. Burrill, S. Melczer, M. Mishna, A Baxter class of a different kind, and other bijective results using tableau sequences ending with a row shape, in *Proceedings of FPSAC 2015*, Discrete Mathematics and Theoretical Computer Science (Association of Discrete Mathematics & Theoretical Computer Science, Nancy, 2015), pp. 369–380
10. S. Burrill, J. Courtiel, E. Fusy, S. Melczer, M. Mishna, Tableau sequences, open diagrams, and Baxter families. Eur. J. Comb. **58**, 144–165 (2016)
11. W.Y.C. Chen, E.Y.P. Deng, R.R.X. Du, R.P. Stanley, C.H. Yan, Crossings and nestings of matchings and partitions. Trans. Am. Math. Soc. **359**(4), 1555–1575 (2007)
12. J. Courtiel, E. Fusy, M. Lepoutre, M. Mishna, Bijections for Weyl chamber walks ending on an axis, using arc diagrams and Schnyder woods. Eur. J. Comb. **69**, 126–142 (2018)
13. D. Denisov, V. Wachtel, Random walks in cones. Ann. Probab. **43**(3), 992–1044 (2015)
14. S. Fomin, Schensted algorithms for dual graded graphs. J. Algebr. Comb. **4**(1), 5–45 (1995)
15. I. Gessel, J. Weinstein, H.S. Wilf, Lattice walks in $\mathbf{Z}^d$ and permutations with no long ascending subsequences. Electron. J. Comb. **5**, Research Paper 2, 11 (1998)
16. I.M. Gessel, Symmetric functions and P-recursiveness. J. Comb. Theory Ser. A **53**(2), 257–285 (1990)
17. I.M. Gessel, D. Zeilberger, Random walk in a Weyl chamber. Proc. Am. Math. Soc. **115**(1), 27–31 (1992)
18. J.B. Gil, P.R.W. McNamara, J.O. Tirrell, M.D. Weiner, From Dyck paths to standard Young tableaux (2017), arXiv:1708.00513
19. B. Gordon, Notes on plane partitions IV. Multirowed partitions with strict decrease along columns, in *Combinatorics, Proceedings of symposia in pure mathematics, University of California, Los Angeles, California, 1968*, vol. XIX, (American Mathematical Society, Providence, 1971), pp. 91–100
20. B. Gordon, L. Houten, Notes on plane partitions I, II. J. Comb. Theory **4**(72–80), 81–99 (1968)
21. B. Gordon, L. Houten, Notes on plane partitions III. Duke Math. J. **36**, 801–824 (1969)
22. I.P. Goulden, D.M. Jackson, *Combinatorial Enumeration* (A Wiley-Interscience Publication, Wiley, New York, 1983). With a foreword by Gian-Carlo Rota, Wiley-Interscience Series in Discrete Mathematics
23. D. Gouyou-Beauchamps, Standard Young tableaux of height 4 and 5. Eur. J. Comb. **10**(1), 69–82 (1989)
24. D.J. Grabiner, Asymptotics for the distributions of subtableaux in Young and up-down tableaux. Electron. J. Comb. **11**(2), Research Paper 29, 22 (2004/2006)
25. D.J. Grabiner, P. Magyar, Random walks in Weyl chambers and the decomposition of tensor powers. J. Algebr. Comb. **2**(3), 239–260 (1993)
26. C. Krattenthaler, Bijections between oscillating tableaux and (semi)standard tableaux via growth diagrams. J. Comb. Theory Ser. A **144**, 277–291 (2016)
27. L. Lipshitz, The diagonal of a D-finite power series is D-finite. J. Algebra **113**(2), 373–378 (1988)
28. S. Okada, Pieri rules for classical groups and equinumeration between generalized oscillating tableaux and semistandard tableaux. Electron. J. Comb. **23**(4), Paper 4.43, 27 (2016)

29. A. Regev, Asymptotic values for degrees associated with strips of young diagrams. Adv. Math. **41**(2), 115–136 (1981)
30. B.E. Sagan, The ubiquitous Young tableau, *Invariant Theory and Tableaux (Minneapolis, MN, 1988)*, vol. 19, The IMA Volumes in Mathematics and its Applications (Springer, New York, 1990), pp. 262–298
31. B.E. Sagan, *The Symmetric Group*, vol. 203, 2nd edn., Graduate Texts in Mathematics (Springer, New York, 2001). Representations, combinatorial algorithms, and aymmetric functions
32. E. Sen-Peng, Skew-standard tableaux with three rows. Adv. Appl. Math. **45**(4), 463–469 (2010)
33. E. Sen-Peng, F. Tung-Shan, J.T. Hou, T.-W. Hsu, Standard Young tableaux and colored Motzkin paths. J. Comb. Theory Ser. B **120**(7), 1786–1803 (2013)
34. R.P. Stanley, Increasing and decreasing subsequences and their variants, in *International Congress of Mathematicians*, vol. I (European Mathematical Society, Zürich, 2007), pp. 545–579
35. R.P. Stanley, Differentiably finite power series. Eur. J. Comb. **1**(2), 175–188 (1980)
36. S. Sundaram, The Cauchy identity for $\mathrm{Sp}(2n)$. J. Comb. Theory Ser. A **53**(2), 209–238 (1990)
37. V.V. Tewari, Kronecker coefficients for some near-rectangular partitions. J. Algebra **429**, 287–317 (2015)
38. G. Viennot, Une forme géométrique de la correspondance de robinson-schensted, in *Combinatoire et repreésentation du groupe symétrique*, vol. 579, Lecture Notes in Mathematics, ed. by D. Foata (Springer, Berlin, 1978), pp. 29–58
39. H.S. Wilf, The computer-aided discovery of a theorem about Young tableaux. J. Symb. Comput. **20**(5–6), 731–735 (1995). Symbolic computation in combinatorics Δ_1 (Ithaca, NY, 1993)
40. H.S. Wilf, Ascending subsequences of permutations and the shapes of tableaux. J. Comb. Theory Ser. A **60**(1), 155–157 (1992)
41. M.C. Wilson, R. Pemantle, *Analytic Combinatorics in Several Variables*, Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, 2013)
42. G. Xin, Determinant formulas relating to tableaux of bounded height. Adv. Appl. Math. **45**(2), 197–211 (2010)
43. D. Zeilberger, The number of ways of walking in $x_1 \geq \cdots \geq x_k \geq 0$ for n days, starting and ending at the origin, where at each day you may either stay in place or move one unit in any direction, equals the number of n-cell standard young tableaux with $\leq 2k+1$ rows (2007). http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimPDF/lazy.pdf

A Tale of Centrally Symmetric Polytopes and Spheres

Isabella Novik

Abstract This paper is a survey of recent advances as well as open problems in the study of face numbers of centrally symmetric simplicial polytopes and spheres. The topics discussed range from neighborliness of centrally symmetric polytopes and the Upper Bound Theorem for centrally symmetric simplicial spheres to the Generalized Lower Bound Theorem for centrally symmetric simplicial polytopes and the lower bound conjecture for centrally symmetric simplicial spheres and manifolds.

1 Introduction

The goal of this paper is to survey recent results related to face numbers of centrally symmetric simplicial polytopes and spheres. To put things into perspective, we start by discussing simplicial polytopes and spheres without a symmetry assumption. The classical theorems of Barnette [10] and McMullen [48], known as the Lower Bound Theorem (LBT, for short) and the Upper Bound Theorem (or UBT), assert that among all d-dimensional simplicial polytopes with n vertices a stacked polytope simultaneously minimizes all the face numbers while the cyclic polytope simultaneously maximizes all the face numbers. Furthermore, the same results continue to hold in the generality of $(d-1)$-dimensional simplicial spheres (see [9, 73]). It is also worth mentioning that in the class of simplicial spheres of dimension $d-1 \geq 3$ with n vertices, the (boundary complexes of the) stacked polytopes are the only minimizers [37]. On the other hand, the maximizers are precisely the $\lfloor d/2 \rfloor$-neighborly

Research is partially supported by NSF grants DMS-1361423 and DMS-1664865, and by Robert R. and Elaine F. Phelps Professorship in Mathematics. This material is based on work supported by the National Science Foundation under Grant No. DMS-1440140 while the author was in residence at the Mathematical Sciences Research Institute in Berkeley CA, during the Fall 2017 semester.

I. Novik
Department of Mathematics, University of Washington,
Seattle, WA 98195-4350, USA
e-mail: novik@math.washington.edu

H. Barcelo et al. (eds.), *Recent Trends in Algebraic Combinatorics*,
Association for Women in Mathematics Series 16,
https://doi.org/10.1007/978-3-030-05141-9_9

spheres—a set that, in addition to the (boundary complex of the) cyclic polytope, includes many other simplicial spheres, see, for instance, a recent paper [63].

In fact, much more is known: the celebrated g-theorem of Billera–Lee [15] and Stanley [74] provides a complete characterization of all possible f-vectors of simplicial polytopes. The equally celebrated g-conjecture posits that the same characterization is valid for the f-vectors of simplicial spheres. (The f-vector encodes the number of faces of each dimension of a simplicial complex.) In other words, at least conjecturally, the f-vector cannot differentiate between simplicial polytopes and simplicial spheres.

Let us now restrict our world and consider all *centrally symmetric* (cs, for short) simplicial polytopes and all *centrally symmetric* simplicial spheres. What is known about the f-vectors of these objects? Are there cs analogs of the LBT and UBT, and ultimately of the g-theorem and g-conjecture, respectively? Surprisingly very little is known, and not for the lack of effort. There is an analog of the LBT for cs simplicial polytopes established by Stanley [75] as well as a characterization of minimizers [41], and while it is plausible that the same results continue to hold for all cs simplicial spheres, the proofs remain elusive.

Still, the most mysterious and fascinating side of the story comes from trying to understand the upper bounds: in contrast with the situation for simplicial polytopes and spheres without a symmetry assumption, the existing upper bound-type results and conjectures indicate striking differences between f-vectors of cs polytopes and those of cs spheres. In other words, cs spheres and polytopes do not look alike f-wise, even though their non-cs counterparts do! For instance, the (appropriately defined) neighborliness of cs polytopes is quite restrictive [46], and a cs d-dimensional polytope with more than 2^d vertices cannot be even 2-neighborly; yet, by a result of Jockusch [36], there exist cs 2-neighborly simplicial spheres of dimension 3 with any even number $n \geq 8$ of vertices. Thus, for a sufficiently large n, the maximum number of edges that a cs 4-dimensional polytope with n vertices can have *differs* from the maximum number of edges that a cs simplicial sphere of dimension 3 with n vertices can have; moreover, the former quantity (or even its asymptotic behavior) is unknown at present. This indicates how very far we are from even posing a plausible (sharp) upper bound conjecture for cs d-dimensional polytopes with $d \geq 4$. At the same time, Adin [2] and Stanley (unpublished) provided upper bounds on face numbers of cs simplicial spheres of dimension $d-1$ with n vertices; these bounds are attained by a cs $\lfloor d/2 \rfloor$-neighborly sphere of dimension $d-1$ with n vertices *assuming* such a sphere exists.

Aside from being of intrinsic interest, additional motivation to better understand the f-vectors of cs polytopes arises from the recently discovered tantalizing connections (initiated by Donoho and his collaborators, see, for instance, [22, 23]) between cs polytopes with many faces and seemingly unrelated areas of error-correcting codes and sparse signal reconstruction. Furthermore, any cs convex body in $\mathbb{R}^d$ is the unit ball of a certain norm on $\mathbb{R}^d$. As a result, methods used in the study of face numbers of cs complexes, at present, involve not only commutative algebra (via the study of associated Stanley–Reisner rings) but also a wealth of techniques from geometric analysis. Yet, many problems remain notoriously difficult.

The goal of this paper is thus to survey a few of the known results on f-vectors of cs simplicial polytopes and spheres, showcase several existing techniques, and, most importantly, present many open problems. It is our hope that collecting such problems in one place will catalyze progress in this fascinating field. For a quick preview of what is known and what is not, see the following table.

	non-cs			**cs**		
	simplicial polytopes	simplicial spheres	same?	simplicial polytope	simplicial spheres	same?
UBT	√	√	yes	no plausible conjecture	known bounds conjecturally sharp	no
LBT	√	√	yes	√	conjecture	conjecturally yes
GLBT	√	conjecture	conjecturally yes	√	conjecture	conjecturally yes

The rest of the paper is structured as follows. In Sect. 2, we set up notation and recall basic definitions pertaining to simplicial polytopes and spheres. Section 3 is devoted to neighborliness of cs polytopes. This leads to discussion of upper bound-type results and problems on face numbers of cs polytopes, see Sect. 4. Section 5 deals with neighborliness and upper bound-type results for cs simplicial spheres. Section 6 takes us into the algebraic side of the story: there we present a quick review of Stanley–Reisner rings—the major algebraic tool in the study of face numbers, sketch the proof of the classical UBT along with Adin–Stanley's variant of this result for cs spheres, and prepare the ground for the following sections. Sections 7 and 8 are concerned with the lower bound-type results and conjectures. Specifically, in Sect. 7 we discuss a cs analog of the Generalized LBT for cs simplicial polytopes—a part of the story that is most well understood, while in Sect. 8 we consider a natural conjectural cs analog of the LBT for spheres, manifolds, and pseudomanifolds. We close with a few concluding remarks in Sect. 9.

2 Preliminaries

We start with outlining basic definitions and notation we will use throughout the paper. A *polytope* is the convex hull of a set of finitely many points in $\mathbb{R}^d$. One example is the (geometric) *d-simplex* defined as the convex hull of an arbitrary set of $d+1$ affinely independent points in $\mathbb{R}^d$. A (proper) *face* of any convex body K (e.g., a polytope) is the intersection of K with a supporting affine hyperplane; see, for example, Chapter II of [11]. A polytope P is called *simplicial* if all of its (proper) faces are simplices. The *dimension* of a polytope P is the dimension of the affine hull of P. We refer to d-dimensional polytopes as d-polytopes and to i-dimensional faces as i-faces.

A polytope $P \subset \mathbb{R}^d$ is *centrally symmetric* (*cs*, for short) if $P = -P$; that is, $x \in P$ if and only if $-x \in P$. An important example of a cs polytope is the d-dimensional *cross-polytope* $\mathcal{C}_d^* = \mathrm{conv}(\pm p_1, \pm p_2, \ldots, \pm p_d)$, where $p_1, p_2, \ldots, p_d$ are points in $\mathbb{R}^d$ whose position vectors form a basis for $\mathbb{R}^d$. If these position vectors form the standard basis of $\mathbb{R}^d$, denoted $\{e_1, \ldots, e_d\}$, then the resulting polytope is the unit ball of the ℓ^1-norm; we refer to this particular instance of the cross-polytope as the *regular* cross-polytope.

A *simplicial complex* Δ on a (finite) vertex set $V = V(\Delta)$ is a collection of subsets of V that is closed under inclusion; an example is the (abstract) simplex on V, $\overline{V} := \{F : F \subseteq V\}$. The elements of Δ are called *faces*, and the maximal under inclusion faces are called *facets*. The *dimension* of a face $F \in \Delta$ is $\dim F = |F| - 1$, and the dimension of Δ is $\dim \Delta := \max\{\dim F \ : \ F \in \Delta\}$. The *k-skeleton* of Δ, $\mathrm{Skel}_k(\Delta)$, is a subcomplex of Δ consisting of all faces of dimension $\leq k$. The *f-vector* of a simplicial complex Δ is $f(\Delta) := (f_{-1}(\Delta), f_0(\Delta), \ldots, f_{\dim \Delta}(\Delta))$, where $f_i = f_i(\Delta)$ denotes the number of i-faces of Δ; the numbers f_i are called the *f-numbers* of Δ.

Each simplicial complex Δ admits a geometric realization $\|\Delta\|$ that contains a geometric i-simplex for each i-face of Δ. A simplicial complex Δ is a *simplicial sphere* (*simplicial ball*, or *simplicial manifold*, respectively) if $\|\Delta\|$ is homeomorphic to a sphere (ball, or closed manifold, respectively). If P is a simplicial d-polytope, then the empty set along with the collection of the vertex sets of all the (proper) faces of P is a simplicial sphere of dimension $d-1$ called the *boundary complex* of P; it is denoted by ∂P. While it follows from Steinitz' theorem that every simplicial 2-sphere can be realized as the boundary complex of a simplicial 3-polytope, for $d \geq 4$, there are many more combinatorial types of simplicial $(d-1)$-spheres than those of boundary complexes of simplicial d-polytopes, see [38, 57, 65].

A simplicial complex Δ is *centrally symmetric* (or cs) if it is equipped with a simplicial *involution* $\phi : \Delta \to \Delta$ such that for every non-empty face $F \in \Delta$, $\phi(F) \neq F$. We refer to F and $\phi(F)$ as *antipodal* faces. For instance, the boundary complex of any cs simplicial polytope P is a cs simplicial sphere with the involution ϕ induced by the map $\phi(v) = -v$ on the vertices of P.

A simplicial complex Δ is *k-neighborly* if every set of k of its vertices forms a face of Δ. Equivalently, a simplicial complex Δ with n vertices is k-neighborly if its $(k-1)$-skeleton coincides with the $(k-1)$-skeleton of the $(n-1)$-simplex. Since in a cs complex, a vertex and its antipode can never form an edge, this definition of neighborliness requires a suitable adjustment for cs complexes. We say that a cs simplicial complex Δ is *k-neighborly* if every set of k of its vertices, no two of which are antipodes, forms a face of Δ. Equivalently, a cs simplicial complex Δ on $2m$ vertices is k-neighborly if its $(k-1)$-skeleton coincides with the $(k-1)$-skeleton of $\partial\mathcal{C}_m^*$. In particular, if Δ is a k-neighborly simplicial complex, then $f_{j-1}(\Delta) = \binom{f_0(\Delta)}{j}$ for all $j \leq k$, while if Δ is a cs k-neighborly simplicial complex, then $f_{j-1}(\Delta) = 2^j\binom{f_0(\Delta)/2}{j}$ for all $j \leq k$.

It is worth mentioning that similar definitions apply to *general* (i.e., not necessarily simplicial) polytopes. Specifically, a polytope P is k-neighborly if every set of k of

its vertices forms the vertex set of a face of P; a cs polytope P is k-neighborly if every set of k of its vertices, no two of which are antipodes, forms the vertex set of a face of P. In the next two sections, we work with general cs polytopes.

3 How Neighborly Can a cs Polytope Be?

Our story begins with the *cyclic polytope*, $C_d(n)$, which is defined as the convex hull of $n \geq d+1$ distinct points on the *moment curve* $\{(t, t^2, \dots, t^d) \in \mathbb{R}^d \,:\, t \in \mathbb{R}\}$ or on the *trigonometric moment curve* $\{(\cos t, \sin t, \cos 2t, \sin 2t, \dots, \cos kt, \sin kt) \in \mathbb{R}^{2k} \,:\, t \in \mathbb{R}\}$, assuming $d = 2k$. Both types of cyclic polytopes were investigated by Carathéodory [19] and later by Gale [27], who, in particular, showed that the two types are combinatorially equivalent (assuming d is even) and independent of the choice of points. In fact, the two types are projectively equivalent, see [84, Exercise 2.21]. These polytopes were also rediscovered by Motzkin [32, 53] and many others; see books [11, 84] for more information on these amazing polytopes. The properties of the cyclic polytope $C_d(n)$ that will be important for us here are: it is a (non-cs) simplicial d-polytope with n vertices; furthermore, it is $\lfloor d/2 \rfloor$-neighborly for all $n \geq d+1$.

The existence of cyclic polytopes motivated several questions, among them: do there exist cs d-polytopes (apart from the cross-polytope) that are $\lfloor d/2 \rfloor$-neighborly or at least "highly" neighborly? This section discusses the state-of-affairs triggered by this question.

It became apparent from the very beginning that the answer is likely to be both interesting and complicated: several works from the sixties indicated that in contrast to the general case, the neighborliness of cs polytopes might be rather restricted. Specifically, Grünbaum showed in 1967 [31, p. 116] that while there exists a cs 2-neighborly 4-polytope with 10 vertices, no cs 4-polytope with 12 or more vertices can be 2-neighborly. This observation was extended by McMullen and Shephard [51] who proved that while there exists a cs $\lfloor d/2 \rfloor$-neighborly d-polytope with $2(d+1)$ vertices, a cs d-polytope with at least $2(d+2)$ vertices cannot be more than $\lfloor (d+1)/3 \rfloor$-neighborly. A cs $\lfloor d/2 \rfloor$-neighborly d-polytope with $2(d+1)$ vertices is easy to construct: for instance, $\operatorname{conv}\left(\pm e_1, \pm e_2, \dots, \pm e_d, \pm \sum_{i=1}^d e_i\right)$ does the job. On the other hand, to show that a cs d-polytope with $2(d+2)$ vertices can only be $\lfloor (d+1)/3 \rfloor$-neighborly and to construct cs polytopes achieving this bound, McMullen and Shephard [51] introduced and studied cs transformations of cs polytopes—a cs analog of celebrated Gale transforms. (See [31, Section 5.4] and [84, Chapter 6] for a detailed description of Gale transforms and their applications.)

Let $k(d, n)$ denote the largest integer k such that there exists a cs d-polytope with $2(n+d)$ vertices that is k-neighborly. Influenced by their result, McMullen and Shephard [51] conjectured that, in fact, $k(d, n) \leq \lfloor (d+n-1)/(n+1) \rfloor$ for all $n \geq 3$. Their conjecture was subsequently refuted by Halsey [33] and then by Schneider [69], but only for $d \gg n$. Namely, Schneider showed that

$$\liminf_{d\to\infty} \frac{k(d,3)}{d+3} \geq 1 - 2^{-1/2} \quad \text{and} \quad \liminf_{d\to\infty} \frac{k(d,n)}{d+n} \geq 0.2390 \text{ for all } n.$$

On the other hand, Burton [18] proved that a cs d-polytope with a sufficiently large number of vertices ($\approx (d/2)^{d/2}$) indeed cannot be even 2-neighborly. Burton's proof is surprisingly short and simple: it relies on John's ellipsoid theorem, see for instance [8, Chapter 3], along with a quantitative version of the observation that any sufficiently large finite subset of the Euclidean unit sphere contains two points that are very close to each other.

To the best of the author's knowledge, McMullen–Shephard's values $k(d,1) = \lfloor \frac{d}{2} \rfloor$ and $k(d,2) = \lfloor \frac{d+1}{3} \rfloor$ (see [51]) remain the *only* known exact values of $k(d,n)$ for all d. In particular, $k(4,1) = 2$ and $k(4,2) = 1$, while $k(5,1) = k(5,2) = 2$. Furthermore, the main result of a very recent paper [60] implies that $k(5,3) = k(5,4) = 2$, but it appears unknown whether $k(5,5)$ equals 2 or 1. On the other hand, the asymptotics of $k(d,n)$ is now well understood:

Theorem 3.1. *There exist absolute constants $C_1, C_2 > 0$ independent of d and n such that*

$$\frac{C_1 d}{1 + \log \frac{n+d}{d}} \leq k(d,n) \leq 1 + \frac{C_2 d}{1 + \log \frac{n+d}{d}}.$$

Theorem 3.1 is due to Linial and Novik [46]. A dual version of the lower bound part of this theorem was also independently established by Rudelson and Vershynin [67].

Two extreme, and hence particularly interesting, cases of Theorem 3.1 deserve a special mention: the case of $k(d,n)$ proportional to d and the case of $k(d,n) = 1$. Donoho [22] proved that there exists $\rho > 0$ such that for large d, the orthogonal projection of the $2d$-dimensional regular cross-polytope onto a d-dimensional subspace of $\mathbb{R}^{2d}$, chosen uniformly at random, is at least $\lfloor \rho d \rfloor$-neighborly with high probability and provided numerical evidence that $\rho \geq 0.089$. In other words, for large d, $k(d,d) \geq \rho d$. (The estimates from [46] guarantee that $\rho \geq 1/400$.) As for the other extreme, Theorem 3.1 shows that the largest number of vertices in a cs 2-neighborly d-polytope is $e^{\Theta(d)}$. In fact, the following more precise result is known, see [46, Theorem 1.2] and [12, Theorem 3.2].

Theorem 3.2. *A cs 2-neighborly d-polytope has at most 2^d vertices, i.e., $k(d, 2^{d-1} + 1 - d) = 1$. On the other hand, for any even $d \geq 6$, there exists a cs 2-neighborly d-polytope with $2(\sqrt{3}^d/3 - 1)$ vertices.*

While this chapter awaited its publication, the author proved in [60] that for every $d \geq 2$, there exists a cs 2-neighborly d-polytope with $2^{d-1} + 2$ vertices.

Our discussion of $k(d,n)$ is summarized in a table below.

We devote the rest of this section to pointing out some of the main ideas used in the proofs. In particular, the proof of the fact that a cs 2-neighborly d-polytope has at most 2^d vertices is so short that we cannot avoid the temptation to provide it here. For a polytope $P \subset \mathbb{R}^d$ and a vector $a \in \mathbb{R}^d$, define $P_a := P + a$ to be the translation of P by a, where "+" denotes the Minkowski addition. The first step in the proof is the following simple observation from [14]:

n	$k(d,n)$
1	$\lfloor \frac{d}{2} \rfloor$
2	$\lfloor \frac{d+1}{3} \rfloor$
d	proportional to d
$\leq 2^{d-2}+1-d$	≥ 2
$\geq 2^{d-1}+1-d$	1

Lemma 3.3. *Let $P \subset \mathbb{R}^d$ be a cs d-polytope, and let u and v be vertices of P, so that $-v$ is also a vertex of P. If the polytopes P_u and P_v have intersecting interiors then the vertices u and $-v$ are not connected by an edge. Consequently, if P is a cs 2-neighborly polytope with vertex set V, then the polytopes $\{P_v : v \in V\}$ have pairwise disjoint interiors.*

Proof: The assumption that $\operatorname{int}(P_u) \cap \operatorname{int}(P_v) \neq \emptyset$ implies that there exist $x, y \in \operatorname{int}(P)$ such that $x + u = y + v$, or equivalently, that $(y - x)/2 = (u - v)/2$. Since P is centrally symmetric, and $x, y \in \operatorname{int}(P)$, the point $q := (y - x)/2$ is an interior point of P. As q is also the barycenter of the line segment connecting u and $-v$, this line segment is not an edge of P. □

The rest of the proof that a cs 2-neighborly d-polytope has at most 2^d vertices utilizes a volume trick that goes back to Danzer and Grünbaum [21]. If $P \subset \mathbb{R}^d$ is a cs 2-neighborly d-polytope with vertex set V, then by Lemma 3.3, the polytopes $\{P_v : v \in V\}$ have pairwise disjoint interiors. Therefore,

$$\operatorname{Vol}\Big(\bigcup_{v\in V} P_v\Big) = \sum_{v\in V} \operatorname{Vol}(P_v) = |V| \cdot \operatorname{Vol}(P).$$

On the other hand, since for $v \in V$, $P_v = P + v \subset 2P$, it follows that $\bigcup_{v\in V} P_v \subseteq 2P$, and so

$$\operatorname{Vol}\Big(\bigcup_{v\in V} P_v\Big) \leq \operatorname{Vol}(2P) = 2^d \cdot \operatorname{Vol}(P).$$

Comparing these two equations yields $|V| \leq 2^d$, as desired.

The proof of the upper bound part of Theorem 3.1 relies on a more intricate application of the same volume trick. Let P be a cs k-neighborly d-polytope, where $k = 2s$ for some integer s. We say that a family $\mathcal{F}$ of $(s-1)$-faces of P is *good* if every two elements $F \neq G$ of $\mathcal{F}$ satisfy the following conditions: F and G share at most $s/2$ vertices, while F and $-G$ have no common vertices. To obtain an upper bound on the size of the vertex set V of P in terms of d and k, one first observes that if $\mathcal{F}$ is a good family, $F \neq G$ are in $\mathcal{F}$, and b_F and b_G are the barycenters of F and G, then the polytopes $P + 2b_F$ and $P + 2b_G$ have disjoint interiors (cf. Lemma 3.3). One then uses a simple counting argument to show that there is a relatively large (in terms of d, s, and $|V|$) good family $\mathcal{F}$. Since the translates $\{P + 2b_F : F \in \mathcal{F}\}$ of P have pairwise disjoint interiors and are all contained in $3P$, the volume trick yields

a desired upper bound on $|\mathcal{F}|$, and hence also on $|V|$; see the proof of Theorem 1.1 in [46] for more details.

The proof of the lower bound in Theorem 3.1 is based on studying the cs transforms of cs polytopes introduced in [51] and on "high-dimensional paradoxes" such as Kašin's theorem [40] and its generalization due to Garnaev and Gluskin [28]. Specifically, Kašin's theorem asserts that there is an absolute constant C (for instance, 32 does the job, see [8, Lecture 4]) such that for every d, $\mathbb{R}^{2d}$ has a d-dimensional subspace, L_d, with the following property: the ratio of the ℓ^2-norm to the ℓ^1-norm of any nonzero vector $x \in L_d$ is in the interval $[\frac{1}{\sqrt{2d}}, \frac{C}{\sqrt{2d}}]$; we refer to such a subspace as a Kašin subspace. Via cs transforms, d-dimensional subspaces of $\mathbb{R}^{2d}$ correspond to cs d-polytopes with $4d$ vertices; furthermore, a careful analysis of cs transforms shows that the polytopes corresponding to Kašin subspaces are k-neighborly with k proportional to d.

More generally, a theorem due to Garnaev and Gluskin [28] quantifies the extent to which an n-dimensional subspace of $\mathbb{R}^{n+d}$ can be "almost Euclidean" (meaning that the ratio of the ℓ^2-norm to the ℓ^1-norm of nonzero vectors remains within certain bounds, more precisely, it is $\leq \tilde{C}\sqrt{\frac{1+\log((n+d)/d)}{d}}$ for some absolute constant $\tilde{C}$). Via cs transforms, n-dimensional subspaces of $\mathbb{R}^{n+d}$ correspond to cs d-polytopes with $2(n+d)$ vertices, and the "almost Euclidean" subspaces correspond to cs polytopes with neighborliness given by the lower bound in Theorem 3.1, see [46] for technical details.

The proof of the Garnaev–Gluskin result and hence also of the lower bound part of Theorem 3.1 is probabilistic in nature: it does not give an explicit construction of neighborly cs polytopes, but rather shows that they form a set of positive probability in a certain probability space. Indeed, it is an interesting open question to find an explicit construction for highly neighborly cs polytopes that meet the lower bound in Theorem 3.1. We discuss some known explicit constructions (for instance that of a cs 2-neighborly d polytope with $\approx \sqrt{3}^d$ vertices) in the next section. It would also be extremely interesting to shed some light on the exact values of $k(d, n)$:

Problem 3.4. *Determine the exact values of $k(d, n)$, or at least find the value of $n_0(d) := \min\{n : k(d, n) = 1\}$, that is, find the number n starting from which a cs d-polytope with $2(d + n)$ vertices cannot be even 2-neighborly.*

4 Toward an Upper Bound Theorem for cs Polytopes

The fame of the cyclic polytope comes from the Upper Bound Theorem (UBT, for short) conjectured by Motzkin [53] and proved by McMullen [48]. It asserts that among all d-polytopes with n vertices, the cyclic polytope $C_d(n)$ simultaneously maximizes all the face numbers. The goal of this section is to summarize several upper bound-type results and problems for cs polytopes motivated by the UBT.

What is the maximum number of k-faces that a *centrally symmetric* d-polytope with N vertices can have? While our discussion in the previous section suggests that at present we are very far from even posing a plausible conjecture, certain asymptotic results on the maximum possible number of edges are known. Specifically, the following generalization of Theorem 3.2 holds. This result is in sharp contrast with the fact that $f_1(C_d(n)) = \binom{n}{2}$ as long as $d \geq 4$.

Theorem 4.1. *Let $d \geq 4$. If $P \subset \mathbb{R}^d$ is a cs d-polytope on N vertices, then*

$$f_1(P) \leq (1 - 2^{-d})\frac{N^2}{2}.$$

On the other hand, there exist cs d-polytopes with N vertices (for an arbitrarily large N) and at least $\left(1 - 3^{-\lfloor d/2-1 \rfloor}\right)\binom{N}{2} \approx (1 - \sqrt{3}^{-d}) \cdot \frac{N^2}{2}$ edges.

The first part of Theorem 4.1 was established by Barvinok and Novik in [14, Proposition 2.1]; its proof relies on an extension of the argument discussed in the previous section to obtain an upper bound on the number of vertices that a cs 2-neighborly d-polytope can have and more specifically on Lemma 3.3. The idea, very roughly, is as follows: each of the N vertex translates of P, P_u for $u \in V$, has the same volume as P, and all of them are contained in the polytope $2P$, whose volume is $2^d \mathrm{Vol}\,(P)$. Hence, on average, an interior point of $2P$ belongs to $N/2^d$ (out of N) sets $\mathrm{int}\,(P_u)$. Therefore, on average, $\mathrm{int}\,(P_u)$ intersects with the interiors of at least $N/2^d - 1$ other vertex translates of P. Consequently, by Lemma 3.3, the average degree of a vertex of P in the graph of P is $\leq N(1 - 2^{-d})$. This yields the desired upper bound on the number of edges of P.

The second part of Theorem 4.1 is due to Barvinok, Lee, and Novik [12]. The proof is based on an explicit construction whose origins can be traced to work of Smilansky [70]. To discuss this part, we start by recalling that the cyclic polytope $C_d(n)$ is defined as the convex hull of n points on the moment curve or, if $d = 2k$ is even, on the trigonometric moment curve $\mathcal{T}_k : \mathbb{R} \to \mathbb{R}^{2k}$, where

$$\mathcal{T}_k(t) = (\cos t, \sin t, \cos 2t, \sin 2t, \ldots, \cos kt, \sin kt).$$

In an attempt to come up with a cs analog of the cyclic polytope, Smilansky [70] (in the case of $k = 2$), and Barvinok and Novik [14] (in the case of arbitrary k) considered the *symmetric moment curve*, $U_k : \mathbb{R} \to \mathbb{R}^{2k}$, defined by

$$U_k(t) = (\cos t, \sin t, \cos 3t, \sin 3t, \ldots, \cos(2k-1)t, \sin(2k-1)t)\,.$$

Since $U_k(t) = U_k(t + 2\pi)$ for all $t \in \mathbb{R}$, from this point on, we think of U_k as defined on the unit circle $\mathbb{S} = \mathbb{R}/2\pi\mathbb{Z}$. The name *symmetric moment curve* is explained by an observation that for all $t \in \mathbb{S}$, t and $t + \pi$ form a pair of opposite points and $U_k(t + \pi) = -U_k(t)$. A *bicyclic polytope* is then defined as the convex hull of $\{U_k(t) : t \in X\}$, where $X \subset \mathbb{S}$ is a finite subset of $\mathbb{S}$; we will also assume that X is a *cs* subset of $\mathbb{S}$.

The papers [70] (in the case of $k = 2$), and [14] along with [79] (in the case of arbitrary k) study the edges of bicyclic polytopes. In particular, when $k = 2$, the following result established in [70] (see also [14]) holds. Recall that a face of a convex body K is the intersection of K with a supporting hyperplane.

Theorem 4.2. *Let $\Gamma \subset \mathbb{S}$ be an open arc of length $2\pi/3$ and let $t_1, t_2 \in \Gamma$. Then the line segment* $\mathrm{conv}(U_2(t_1), U_2(t_2))$ *is an edge of the 4-dimensional convex body* $\mathcal{B}_2 := \mathrm{conv}(U_2(t) : t \in \mathbb{S})$.

One immediate consequence is

Corollary 4.3. *Let $X = \{0, \pi/2, \pi, 3\pi/2\} \subset \mathbb{S}$, let $s \geq 2$ be an integer, and let X_s be a cs subset of $\mathbb{S}$ obtained from X by replacing each $\tau \in X$ with a cluster of s points all of which lie on a small arc containing τ. Then the polytope $Q_s :=$* $\mathrm{conv}(U_2(t) : t \in X_s)$ *is a cs 4-polytope that has $N := 4s$ vertices and at least $\frac{1}{2} \cdot N(\frac{3}{4}N - 1) \approx \frac{3}{4}\binom{N}{2}$ edges.*

Indeed, it follows from Theorem 4.2 that each vertex of Q_s is connected by an edge to all other vertices of Q_s except possibly those coming from the opposite cluster, yielding the result.

Denote by $\mathrm{fmax}\,(d, N; k-1)$ the maximum possible number of $(k-1)$-faces that a cs d-polytope with N vertices can have. From the above discussion, we infer that

$$\frac{3}{4} \cdot \frac{N^2}{2} - O(N) \leq \mathrm{fmax}\,(4, N; 1) \leq \frac{15}{16} \cdot \frac{N^2}{2}.$$

These are currently the *best*-known bounds on the maximum possible number of edges that a cs 4-polytope with N vertices can have.

Perhaps somewhat surprisingly, for $k > 2$, bicyclic polytopes do not have a record number of edges. However, the symmetric moment curve is used in the following construction that produces cs polytopes with the largest known to-date number of edges.

Let $m \geq 1$ be an integer. Define $\Phi_m(t) : \mathbb{S} \to \mathbb{R}^{2m+2}$ by

$$\begin{aligned}\Phi_m(t) &:= (U_1(t), U_1(3t), U_1(3^2t), \dots, U_1(3^m(t)) \\ &= (\cos t, \sin t, \cos 3t, \sin 3t, \cos 9t, \sin 9t, \dots, \cos(3^m t), \sin(3^m t)).\end{aligned}$$

Parts 1 and 2 of the following result complete the proofs of Theorems 3.2 and 4.1, respectively.

Theorem 4.4. *Fix integers $m \geq 2$ and $s \geq 2$. Let A_m be a set of $2(3^m - 1)$ equally spaced points on $\mathbb{S}$, and let $A_{m,s}$ be a cs subset of $\mathbb{S}$ obtained from A_m by replacing each $\tau \in A_m$ with a cluster of s points, all of which lie on a very small arc containing τ. Then*

1. *the polytope $P_m :=$* $\mathrm{conv}(\Phi_m(A_m))$ *is a cs 2-neighborly polytope of dimension $2(m+1)$ that has $2(3^m - 1)$ vertices,*

2. *the polytope* $P_{m,s} := \mathrm{conv}(\Phi_m(A_{m,s}))$ *is a cs polytope of dimension* $2(m+1)$ *that has* $N := 2s(3^m - 1)$ *vertices and more than* $\left(1-3^{-m}\right)\binom{N}{2}$ *edges.*

The assumption $m \geq 2$ is only needed to guarantee that the dimension of P_m is exactly $2(m+1)$ rather than $\leq 2(m+1)$. To prove that P_m is cs 2-neighborly and $P_{m,s}$ has many edges for all $m \geq 1$, it is enough to show that each vertex of $P_{m,s}$ is connected by an edge to all other vertices of $P_{m,s}$ except possibly those coming from the opposite cluster. This can be done by induction on m. Since $\Phi_1 = U_2$, the case of $m = 1$ is simply Corollary 4.3. For the inductive step, one relies on some standard facts about faces of polytopes along with an observation that the composition of $\Phi_m : \mathbb{R} \to \mathbb{R}^{2(m+1)}$ with the projection of $\mathbb{R}^{2(m+1)}$ onto $\mathbb{R}^{2m}$ that forgets the first two coordinates is the curve $t \to \Phi_{m-1}(3t)$, while the composition of Φ_m with the projection of $\mathbb{R}^{2(m+1)}$ onto $\mathbb{R}^4$ that forgets all but the first four coordinates is the curve Φ_1; see [12, Section 3] for more details. Finally, to extend the construction of Theorem 4.4 to *odd* dimensions, consider the bipyramid over polytopes P_m and $P_{m,s}$.

To summarize,

$$\left(1-3^{-\lfloor d/2-1\rfloor}\right)\binom{N}{2} \leq \mathrm{fmax}\,(d, N; 1) \leq \left(1-2^{-d}\right)\cdot\frac{N^2}{2}. \tag{4.1}$$

It is, however, worth pointing out that in view of the main result of [60], $\mathrm{fmax}\,(d, N; 1)$ might be closer to the right-hand side of Eq. (4.1) than to the left one.

To extend the above discussion to higher-dimensional faces, we need to look more closely at the curve U_k and its convex hull.

One crucial feature of the convex hull $\mathcal{M}_k = \mathrm{conv}(\mathcal{T}_k(t) : t \in \mathbb{S}) \subset \mathbb{R}^{2k}$ of the trigonometric moment curve is that it is k-neighborly, that is, for any $p \leq k$ distinct points $t_1, \ldots, t_p \in \mathbb{S}$, the convex hull $\mathrm{conv}(\mathcal{T}_k(t_1), \ldots, \mathcal{T}_k(t_p))$ is a face of $\mathcal{M}_k$; see, for example, Chapter II of [11]. While the convex hull of U_k is not k-neighborly, the following theorem, which is the main result of [13], shows that it is *locally* k*-neighborly* (cf. Theorem 4.2).

Theorem 4.5. *For every positive integer k there exists a number* $\frac{\pi}{2} < \alpha_k < \pi$ *such that for an arbitrary open arc* $\Gamma \subset \mathbb{S}$ *of length* α_k *and arbitrary distinct* $p \leq k$ *points* $t_1, \ldots, t_p \in \Gamma$*, the set* $\mathrm{conv}(U_k(t_1), \ldots, U_k(t_p))$ *is a face of* $\mathcal{B}_k := \mathrm{conv}(U_k(t) : t \in \mathbb{S})$.

The gap between the currently known upper and lower bounds on $\mathrm{fmax}\,(d, N; k-1)$ for $k > 2$ is much worse than the gap for $k = 2$ illustrated by Eq. (4.1). Indeed, we have:

Theorem 4.6. *Let* $3 \leq k \leq d/2$*. Then*

$$\left(1-k^2\left(2^{3/20k^2 2^k}\right)^{-d}\right)\binom{N}{k} \leq \mathrm{fmax}\,(d, N; k-1) \leq \left(1-2^{-d}\right)\frac{N}{N-1}\binom{N}{k}.$$

The proof of the upper bound part of this theorem follows easily from the first part of Theorem 4.1 together with (1) the well-known perturbation trick that reduces the

situation to cs *simplicial* polytopes and (2) the standard double-counting argument that relates the number of edges to the number of $(k-1)$-faces in any simplicial complex with N vertices, see [14, Proposition 2.2] for more details.

The proof of the lower bound part can be found in [12, Section 4]. It utilizes a construction that is somewhat along the lines of the construction of Theorem 4.4, but quite a bit more involved: the desired polytope is obtained as the convex hull of a carefully chosen set of points on the curve $\Psi_{k,m} : \mathbb{S} \to \mathbb{R}^{2k(m+1)}$ defined by

$$\Psi_{k,m}(t) := (U_k(t), U_k(3t), U_k(3^2t), \dots, U_k(3^m t)).$$

(Note that Φ_m is essentially $\Psi_{2,m}$: in $\Psi_{2,m}(t)$ every coordinate except for the first two and the last two shows up twice; to obtain Φ_m from $\Psi_{2,m}$ simply leave one copy of each repeated coordinate.) To choose an appropriate set of points one uses the notion of a k-independent family of subsets of $\{1, 2, \dots, m\}$ and a deterministic construction from [25] of a large k-independent family. Finally, to show that the resulting polytope has many $(k-1)$-faces, one relies heavily on the local neighborliness of the convex hull of U_k discussed in Theorem 4.5 along with some standard results on faces of polytopes.

As the discussion above indicates, at present our understanding of the maximum possible number of faces of a given dimension that a cs d-polytope with a given number of vertices can have is rather limited even for $d = 4$. In particular, the following questions are wide open:

Problem 4.7. *Does the limit* $\lim_{N\to\infty} \frac{\mathrm{fmax}\,(d,N;1)}{\binom{N}{2}}$ *exist and, if so, what is its value? Or better yet, what is the actual value of* $\mathrm{fmax}\,(d, N; 1)$*? Similarly, what is* $\lim_{N\to\infty} \frac{\mathrm{fmax}\,(d,N;k-1)}{\binom{N}{k}}$ *for* $2 < k \le d/2$*? Can we, at least, establish better lower and upper bounds than those given in Theorem 4.6?*

In fact, it should be stressed that for $d \ge 6$ and $N \ge 2(d+2)$, we do not even know if in the class of cs d-polytopes with N vertices there is a polytope that simultaneously maximizes all the face numbers. (For $d = 4, 5$, a cs simplicial polytope that maximizes the number of edges automatically maximizes the rest of face numbers.)

5 Toward an Upper Bound Theorem for cs Simplicial Spheres

McMullen's Upper Bound Theorem was extended by Stanley [73] to the class of all simplicial spheres. More precisely, Stanley proved that among all simplicial $(d-1)$-spheres with n vertices, the boundary complex of the cyclic *polytope* $C_d(n)$ simultaneously maximizes all the face numbers. (This result was extended even further to some classes of simplicial manifolds and even certain pseudomanifolds with isolated singularities, see [34, 58, 62]. In fact, already in 1964, Victor Klee [44] proved that the UBT holds for all $(d-1)$-dimensional *Eulerian* complexes that have at least

$O(d^2)$ vertices. Whether the UBT holds for all Eulerian complexes remains an open question.)

The situation with the face numbers of cs spheres versus face numbers of cs polytopes appears to be drastically different. On one hand, as we saw in Sects. 3 and 4, a cs 4-polytope P with $N \geq 12$ vertices cannot be 2-neighborly; furthermore, such a P has at most $\frac{15}{16} \cdot \frac{N^2}{2}$ edges. On the other hand, it is a result of Jockusch [36] that

Theorem 5.1. *For every even number $N \geq 8$, there exists a cs 2-neighborly simplicial sphere J_N of dimension 3 with N vertices. In particular, J_N has $\binom{N}{2} - \frac{N}{2} = \frac{N^2}{2} - N$ edges.*

Jockusch's proof, see [36], is by inductive construction. The initial sphere, J_8, is the boundary complex of the cross-polytope $\mathcal{C}_4^*$ with the involution ϕ induced by the map $\phi(v) = -v$ on the vertex set of $\mathcal{C}_4^*$. The cs sphere (J_{N+2}, ϕ) is obtained from the cs sphere (J_N, ϕ) by the following procedure:

1. Find a subcomplex B_N in J_N such that (i) B_N is a 3-ball, (ii) B_N and $\phi(B_N)$ share no common facets, (iii) every vertex of J_N is a vertex of B_N, and (iv) every edge of B_N lies on the *boundary* of B_N, i.e., B_N has no interior vertices and no interior edges.
2. Let v_{N+1} and v_{N+2} be two new vertices. Cut out the interior of B_N from J_N, and cone the boundary of the resulting hole with v_{N+1}. Similarly, cut out the interior of $\phi(B_N)$ from J_N, and cone the boundary of the resulting hole with v_{N+2}.

Extending ϕ to J_{N+2} by letting $\phi(v_{N+1}) = v_{N+2}$ and $\phi(v_{N+2}) = v_{N+1}$ provides us with a free involution on J_{N+2}. Furthermore, choosing B_N in a way that is specified in Part (1) of the construction guarantees that the cs sphere J_{N+2} is 2-neighborly (provided J_N was 2-neighborly). To allow for this inductive construction, an essential part of Jockusch's proof is devoted to defining B_N in a way that ensures that the resulting simplicial sphere J_{N+2} has a subcomplex B_{N+2} with the same properties.

The *suspension* of a simplicial complex Δ, $\Sigma(\Delta)$, is the join of Δ with the 0-dimensional sphere, that is, $\Sigma(\Delta) = \{F, F \cup \{u_0\}, F \cup \{w_0\} : F \in \Delta\}$, where u_0 and w_0 are two new vertices. Observe that if Δ is cs, then $\Sigma(\Delta)$ is cs (with u_0 and w_0 being antipodes), and if Δ is a $(d-1)$-sphere, then $\Sigma(\Delta)$ is a d-sphere; furthermore, if Δ is cs k-neighborly, then so is $\Sigma(\Delta)$. In particular, for every even $N \geq 8$, $\Sigma(J_N)$ is a cs 2-neighborly simplicial 4-sphere with $N+2$ vertices. Jockusch's results lead to the following problem on higher-dimensional cs spheres.

Problem 5.2. *Let $d > 5$ and let $N \geq 2d$ be an even number. Is there a cs $\lfloor d/2 \rfloor$-neighborly simplicial $(d-1)$-sphere with N vertices?*

The boundary complex of $\mathcal{C}_d^*$ is the unique cs d-neighborly simplicial $(d-1)$-sphere with $2d$ vertices. The boundary complex of $\operatorname{conv}(\pm e_1, \ldots, \pm e_d, \pm \sum_{i=1}^{d} e_i)$ is a cs $\lfloor d/2 \rfloor$-neighborly simplicial $(d-1)$-sphere with $2(d+1)$ vertices [51]; furthermore, in the case of an odd d, [16, Section 6.2] provides a construction of many cs $\lfloor d/2 \rfloor$-neighborly simplicial $(d-1)$-spheres with $2(d+1)$ vertices. Lutz [47, Chapter 4] found (by a computer search) several cs 3-neighborly simplicial 5-spheres

with $16 = 2(6+2)$ vertices; his examples have vertex-transitive cyclic symmetry. Suspensions of Lutz's examples are cs 3-neighborly simplicial 6-spheres with 18 vertices. For all other values of d and N, Problem 5.2 remains wide open. It is also worth pointing out that Pfeifle [64, Chapter 10] investigated possible neighborliness of cs star-shaped spheres—a class of objects that contains all boundary complexes of cs simplicial polytopes and is contained in the class of all cs simplicial spheres. In analogy with McMullen–Shephard's result [51] about cs d-polytopes with $2(d+2)$ vertices, he proved that for all even $d \geq 4$, and for all odd $d \geq 11$, there are no cs $\lfloor d/2 \rfloor$-neighborly star-shaped spheres of dimension $d-1$ with $2(d+2)$ vertices.

One of the main reasons for trying to resolve Problem 5.2 comes from the following cs analog of the Upper Bound Theorem; this result is due to Adin [2] and Stanley (unpublished).

Theorem 5.3. *Among all cs simplicial $(d-1)$-spheres with N vertices, a cs $\lfloor d/2 \rfloor$-neighborly $(d-1)$-sphere with N vertices simultaneously maximizes all the face numbers, assuming such a sphere exists.*

We sketch the proof of Theorem 5.3 in the next section; as in the classic non-cs case, it relies on the theory of Stanley–Reisner rings and on the Dehn–Sommerville relations.

To close this section, we posit a weaker and hence potentially more approachable version, of Problem 5.2. Let Δ be a $(d-1)$-dimensional simplicial complex and F a face of Δ. The *link* of F in Δ is $\mathrm{lk}_\Delta(F) := \{G \in \Delta \ : \ F \cup G \in \Delta, \ F \cap G = \emptyset\}$. (Thus, the link of F describes the local structure of Δ around F.) We say that Δ is *pure* if all facets of Δ have dimension $d-1$; equivalently, for any face $F \in \Delta$, the link of F is $(d-|F|-1)$-dimensional. Furthermore, we say that Δ is *Eulerian* if it is pure and the link of any face $F \in \Delta$ (including the empty face) has the same Euler characteristic as $\mathbb{S}^{d-|F|-1}$—the $(d-|F|-1)$-dimensional sphere. In particular, the class of Eulerian complexes includes all simplicial spheres, all odd-dimensional simplicial manifolds as well as all even-dimensional manifolds whose Euler characteristic is two. Eulerian complexes were introduced by Victor Klee in [43].

Problem 5.4. *Let $d > 5$ and let $N \geq 2d$ be an even number. Is there a cs $\lfloor d/2 \rfloor$-neighborly $(d-1)$-dimensional Eulerian complex with N vertices?*

6 The Algebraic Side of the Story: Stanley–Reisner Rings

We now switch to the algebraic side of the story and present a quick review of the major algebraic tool in the study of face numbers of simplicial complexes—the Stanley–Reisner ring. Along the way, we outline the proof of Theorem 5.3 as well as prepare the ground for our discussion of the Lower Bound Theorems in the next section. The main reference to this material is Stanley's book [77].

Let Δ be a simplicial complex with vertex set V and let $\mathbb{K}$ be an *infinite* field of an arbitrary characteristic. Consider the polynomial ring $S := \mathbb{K}[x_v \ : \ v \in V]$ with

one variable for each vertex in Δ. The *Stanley–Reisner ideal* of Δ is the following squarefree monomial ideal

$$I_\Delta := \left(x_{v_1}x_{v_2}\cdots x_{v_k} \ : \ \{v_1, v_2, \ldots, v_k\} \notin \Delta\right).$$

The *Stanley–Reisner ring* (or face ring) of Δ is the quotient $\mathbb{K}[\Delta] := S/I_\Delta$. Since I_Δ is a monomial ideal, the quotient ring $\mathbb{K}[\Delta]$ is graded by degree. The definition of I_Δ ensures that, as a $\mathbb{K}$-vector space, each graded piece of $\mathbb{K}[\Delta]$, denoted $\mathbb{K}[\Delta]_i$, has a natural basis of monomials whose supports correspond to faces of Δ.

Stanley's and Hochster's insight (independently from each other) in defining this ring [35, 73] was that algebraic properties of $\mathbb{K}[\Delta]$ reflect many combinatorial and topological properties of Δ. For instance, if Δ is $(d-1)$-dimensional, then the Krull dimension of $\mathbb{K}[\Delta]$ is d; in fact, the Hilbert series of $\mathbb{K}[\Delta]$, i.e., $\mathrm{Hilb}(\mathbb{K}[\Delta], t) := \sum_{i=0}^{\infty} \dim_{\mathbb{K}} \mathbb{K}[\Delta]_i \cdot t^i$, is given by

$$\mathrm{Hilb}(\mathbb{K}[\Delta], t) = \sum_{i=0}^{d} \frac{f_{i-1}(\Delta)t^i}{(1-t)^i} = \frac{\sum_{i=0}^{d} f_{i-1}(\Delta)t^i(1-t)^{d-i}}{(1-t)^d}.$$

The last equation leads to the following definition: if Δ is a $(d-1)$-dimensional simplicial complex, then the *h-vector* of Δ, $h(\Delta) = (h_0(\Delta), h_1(\Delta), \ldots, h_d(\Delta))$, is the vector whose entries satisfy $\sum_{i=0}^{d} h_i(\Delta)t^i = \sum_{i=0}^{d} f_{i-1}(\Delta)t^i(1-t)^{d-i}$; equivalently,

$$\sum_{i=0}^{d} h_i(\Delta)t^{d-i} = \sum_{i=0}^{d} f_{i-1}(\Delta)(t-1)^{d-i}. \tag{6.1}$$

In particular, $h_0(\Delta) = 1$, $h_1(\Delta) = f_0(\Delta) - d$, and $h_2(\Delta) = f_1(\Delta) - (d-1)f_0(\Delta) + \binom{d}{2}$.

The following immediate consequences of Eq. (6.1) are worth mentioning: knowing the f-numbers of Δ is equivalent to knowing its h-numbers. Moreover, since the f-numbers are *nonnegative* integer combinations of the h-numbers, any upper/lower bounds on the h-numbers of Δ automatically imply upper/lower bounds on the f-numbers of Δ.

Let Δ be a $(d-1)$-dimensional simplicial complex. A sequence of linear forms, $\theta_1, \theta_2, \ldots, \theta_d \in S$ is called a *linear system of parameters* (or l.s.o.p.) for $\mathbb{K}[\Delta]$ if the ring $\mathbb{K}[\Delta]/(\Theta)$ is a finite-dimensional $\mathbb{K}$-vector space; here $(\Theta) := (\theta_1, \ldots, \theta_d)$. It is well known that if $\mathbb{K}$ is an infinite field, then $\mathbb{K}[\Delta]$ admits an l.s.o.p. An l.s.o.p. for $\mathbb{K}[\Delta]$ is a *regular sequence* if θ_{i+1} is a nonzero divisor on $\mathbb{K}[\Delta]/(\theta_1, \ldots, \theta_i)$ for all $0 \le i < d$. We say that Δ is *Cohen–Macaulay* (over $\mathbb{K}$), or $\mathbb{K}$-*CM* for short, if every l.s.o.p. for $\mathbb{K}[\Delta]$ is a regular sequence.

Assume now that Δ is $\mathbb{K}$-CM, and $\theta_1, \ldots, \theta_d$ is an l.s.o.p. for $\mathbb{K}[\Delta]$. Then θ_1 is a nonzero divisor on $\mathbb{K}[\Delta]$, and so the following sequence of $\mathbb{K}$-vector spaces

$$0 \to \mathbb{K}[\Delta]_{i-1} \xrightarrow{\cdot\theta_1} \mathbb{K}[\Delta]_i \longrightarrow \left(\mathbb{K}[\Delta]/(\theta_1)\right)_i \to 0 \tag{6.2}$$

is exact (for all $i \geq 0$). Thus, $\dim_{\mathbb{K}}(\mathbb{K}[\Delta]/(\theta_1))_i = \dim_{\mathbb{K}} \mathbb{K}[\Delta]_i - \dim_{\mathbb{K}} \mathbb{K}[\Delta]_{i-1}$ for all i. Multiplying by t^i and summing up the resulting equations, we obtain that $\mathrm{Hilb}(\mathbb{K}[\Delta]/(\theta_1), t) = (1-t)\,\mathrm{Hilb}(\mathbb{K}[\Delta], t)$. Iterating this argument for $\theta_2, \ldots, \theta_d$ leads to the following result due to Stanley [73, Section 4].

Theorem 6.1. *Let Δ be a $(d-1)$-dimensional $\mathbb{K}$-CM complex and let $\theta_1, \ldots, \theta_d$ be an l.s.o.p. for $\mathbb{K}[\Delta]$. Then $\mathrm{Hilb}(\mathbb{K}[\Delta]/(\Theta), t) = (1-t)^d\,\mathrm{Hilb}(\mathbb{K}[\Delta], t) = \sum_{i=0}^d h_i(\Delta)t^i$. (In particular, the h-numbers of CM complexes are nonnegative.)*

If $\Delta \subseteq \Gamma$ are simplicial complexes (say, on the same vertex set V), then $I_\Delta \supseteq I_\Gamma$, and so there is a natural surjection $\mathbb{K}[\Gamma] \to \mathbb{K}[\Delta]$. Furthermore, if $\dim \Delta = \dim \Gamma$, then any l.s.o.p. $\theta_1, \ldots, \theta_d$ for $\mathbb{K}[\Gamma]$ is also an l.s.o.p. for $\mathbb{K}[\Delta]$, and the induced graded homomorphism $\mathbb{K}[\Gamma]/(\Theta) \to \mathbb{K}[\Delta]/(\Theta)$ is surjective. This observation together with Theorem 6.1 yields the following special case of [76, Theorem 2.1]:

Corollary 6.2. *Let $\Delta \subseteq \Gamma$ be simplicial complexes. Assume that both Δ and Γ are $\mathbb{K}$-CM and have the same dimension. Then $h_i(\Delta) \leq h_i(\Gamma)$ for all i.*

The reason CM complexes are relevant to our discussion is that, by a result of Reisner [66], Δ is a $\mathbb{K}$-CM complex if and only if Δ is pure and for every face F of Δ (including the empty face), all but the top homology group (computed over $\mathbb{K}$) of the link of F vanish. In particular, all simplicial spheres and balls are CM over any field; furthermore, the j-skeleton of a $\mathbb{K}$-CM complex is $\mathbb{K}$-CM for all j.

Another very important result about simplicial spheres is Dehn–Sommerville relations established by Klee [43]: if Δ is a simplicial $(d-1)$-sphere, then $h_i(\Delta) = h_{d-i}(\Delta)$ for all $0 \leq i \leq d$. (In fact, Klee showed that these relations hold for all Eulerian complexes.)

We are now ready to close this section with a sketch of the proof of Stanley's UBT and of its cs analog—Theorem 5.3. Following the custom, if P is a simplicial polytope, we denote by $h(P)$ the h-vector of the boundary complex of P.

Let Δ be a simplicial $(d-1)$-sphere with vertex set V, $|V| = n$. Let $\overline{V}$ be the simplex on V, let $\Gamma = \mathrm{Skel}_{d-1}(\overline{V})$, and let $C_d(n)$ be the cyclic polytope. Then $\Delta \subseteq \Gamma$ are both CM complexes of the same dimension. Hence, by Corollary 6.2, $h_i(\Delta) \leq h_i(\Gamma)$ for all $0 \leq i \leq d$. Furthermore, since $C_d(n)$ is $\lfloor d/2 \rfloor$-neighborly, $\partial C_d(n)$ and Γ have the same $(\lfloor d/2 \rfloor - 1)$-skeleton, and so $h_i(\Gamma) = h_i(C_d(n))$ for all $0 \leq i \leq d/2$. (Indeed, h_i is determined by $f_0, f_1, \ldots, f_{i-1}$.) We conclude that $h_i(\Delta) \leq h_i(C_d(n))$ for all $0 \leq i \leq d/2$. Dehn–Sommerville relations, applied to Δ and $\partial C_d(n)$, then yield that $h_i(\Delta) \leq h_i(C_d(n))$ also holds for all $d/2 \leq i \leq d$. Thus, $h_i(\Delta) \leq h_i(C_d(n))$ for all $0 \leq i \leq d$, and the Upper Bound Theorem, asserting that $f_j(\Delta) \leq f_j(C_d(n))$ for all $1 \leq j \leq d-1$, follows.

Similarly, if Δ is a cs simplicial $(d-1)$-sphere with $N = 2m$ vertices, then Δ is a full-dimensional subcomplex of $\Gamma := \mathrm{Skel}_{d-1}(\partial \mathcal{C}_m^*)$ (under any bijection from the vertex set of Δ to that of $\partial \mathcal{C}_m^*$ that takes antipodal vertices to antipodal vertices). Further, if S is a cs $\lfloor d/2 \rfloor$-neighborly simplicial $(d-1)$-sphere with N vertices, then $\mathrm{Skel}_{\lfloor d/2 \rfloor - 1}(\Gamma) = \mathrm{Skel}_{\lfloor d/2 \rfloor - 1}(S)$. Applying the same argument as above to Δ, Γ, and S, we conclude that $h_i(\Delta) \leq h_i(S)$ for all $0 \leq i \leq d$, and hence that $f_j(\Delta) \leq f_j(S)$ for all $1 \leq j \leq d-1$. This completes the proof of Theorem 5.3.

Note that whether a cs $\lfloor d/2 \rfloor$-neighborly simplicial $(d-1)$-sphere with $N = 2m$ vertices exists or not, the h-numbers (and hence also f-numbers) such a sphere would have are well defined: $h_i(S) = h_{d-i}(S) = h_i(\mathrm{Skel}_{d-1}(\mathcal{C}_m^*))$ for all $i \leq d/2$. Thus, independently of the existence of such a sphere, Theorem 5.3 provides upper bounds on face numbers of cs simplicial $(d-1)$-spheres with N vertices. However, if a cs $\lfloor d/2 \rfloor$-neighborly simplicial $(d-1)$-sphere with N vertices exists, then these bounds are tight.

Note also that the proof of Theorem 5.3 almost does not use the central symmetry assumption: instead, it relies on a much weaker condition, namely, on the existence of a free involution $\phi : V(\Delta) \to V(\Delta)$ on the vertex set of Δ such that $\{v, \phi(v)\}$ is not an edge of Δ for all $v \in V(\Delta)$. We refer our readers to [59] for more details on the h-vectors of simplicial spheres and manifolds possessing this weaker property and to [17] for a complete characterization of h-vectors of CM complexes with this property.

7 The Generalized Lower Bound Theorem for cs Polytopes

Our ultimate dream is to find a cs analog of the g-theorem for cs simplicial polytopes. While at the moment it is completely out of reach (we do not even have a plausible upper bound conjecture, let alone a complete characterization!), establishing the lower bound-type results is a necessary part of the program. To this end, in this section we discuss a cs analog of the Generalized Lower Bound Theorem for cs simplicial polytopes. We start by reviewing the classical Lower Bound Theorem (LBT, for short) and the Generalized Lower Bound Theorem (GLBT, for short). To state these results, we need a few definitions.

A *triangulation* of a simplicial d-polytope P is a simplicial d-ball B whose boundary, ∂B, coincides with ∂P. A simplicial d-polytope P is called $(r-1)$-*stacked* (for some $1 \leq r \leq d$) if there exists a triangulation B of P such that $\mathrm{Skel}_{d-r}(B) = \mathrm{Skel}_{d-r}(\partial P)$, i.e., all "new" faces of this triangulation have dimension $> d - r$. Note that the simplices are the only 0-stacked polytopes, and that 1-stacked polytopes—usually referred to as *stacked* polytopes—are polytopes that can be obtained from the d-simplex by repeatedly attaching (shallow) d-simplices along facets. In contrast with the cyclic polytopes, two stacked d-polytopes with n vertices may not have the same combinatorial type. However, they do have the same face numbers.

Theorem 7.1. (LBT) *Let P be a simplicial d-polytope with $d \geq 4$. Then $h_1(P) \leq h_2(P)$, with equality if and only if P is stacked.*

Theorem 7.2. (GLBT) *Let P be a simplicial d-polytope. Then*

$$1 = h_0(P) \leq h_1(P) \leq h_2(P) \leq \cdots \leq h_{\lfloor d/2 \rfloor}(P).$$

Furthermore, $h_r(P) = h_{r-1}(P)$ for some $r \leq d/2$ if and only if P is $(r-1)$-stacked.

Since f_{i-1} is a nonnegative linear combination of $h_0, h_1, \ldots, h_i$, one immediate corollary of Theorem 7.1 is that among all simplicial d-polytopes with n vertices, a stacked polytope has the smallest number of edges. In fact, Theorem 7.1 together with a well-known reduction due to McMullen, Perles, and Walkup implies that among all simplicial d-polytopes with n vertices, stacked polytopes simultaneously minimize *all* the face numbers, and that for $d \geq 4$, stacked polytopes are the only minimizers, see [37, Section 5].

The differences between consecutive h-numbers are known as the *g-numbers*: $g_0(P) := 1$ and $g_r(P) := h_r(P) - h_{r-1}(P)$ for $1 \leq r \leq d/2$. Nonnegativity of g_2 for simplicial polytopes of dimension 4 and higher was established by Barnette [10], while Billera and Lee [15] proved that the equality $g_2(P) = 0$ (for $d \geq 4$) holds if and only if P is stacked. The first part of the GLBT was proved by Stanley [74] (a more elementary proof of this part is due to McMullen [49, 50]). The second part was conjectured by McMullen and Walkup, [52] and proved only recently by Murai and Nevo [55].

Since the d-dimensional cross-polytope has the minimal number of faces among all cs simplicial d-polytopes and since $h_r(\mathcal{C}_d^*) = \binom{d}{r} > \binom{d}{r-1} = h_{r-1}(\mathcal{C}_d^*)$ for all $1 \leq r \leq d/2$, one expects that for this class of polytopes, the inequalities $g_r(P) \geq 0$ of the GLBT can be considerably strengthened. Indeed, the following result of Stanley [75] provides a cs version of the inequality part of the GLBT. This result settled an unpublished conjecture by Björner.

Theorem 7.3. *Let P be a cs simplicial d-polytope. Then*

$$g_r(P) \geq \binom{d}{r} - \binom{d}{r-1} = g_r(\mathcal{C}_d^*) \quad \textit{for all } r \leq d/2.$$

Stanley's proof of the first part of Theorem 7.2 is based on the theory of toric varieties associated with polytopes. Here is a very rough sketch of the argument. If $P \subset \mathbb{R}^d$ is a simplicial d-polytope, then a slight perturbation of the vertices of P does not change its combinatorial type, and so we may assume without loss of generality that P has rational vertices; further, by translating P, we may also assume that the origin is in the interior of P. Let V be the vertex set of P. For each $v \in V$, let $(v_1, v_2, \ldots, v_d)$ denote the coordinates of v in $\mathbb{R}^d$, and define $\theta_j := \sum_{v \in V} v_j x_v \in \mathbb{R}[x_v \ : \ v \in V]$ for $j = 1, 2, \ldots, d$. Consider $\mathbb{R}[\partial P]$—the Stanley–Reisner ring of the boundary complex of P over $\mathbb{R}$. Then $\theta_1, \ldots, \theta_d$ is an l.s.o.p. of $\mathbb{R}[\partial P]$ (this follows, for instance, from [77, Theorem III.2.4]), and so by Theorem 6.1, $\dim_{\mathbb{R}} (\mathbb{R}[\partial P]/(\Theta))_i = h_i(P)$. On the other hand, $\mathbb{R}[\partial P]/(\Theta)$ is isomorphic to the singular cohomology ring of the toric variety X_P corresponding to P (see [20, Theorem 10.8] and [26, Section 5.2] for more details). Let $\omega = \sum_{v \in V} x_v$. The toric variety X_P is known to satisfy the hard Lefschetz theorem, which implies that for all $i \leq d/2$, the multiplication map $\cdot\omega : (\mathbb{R}[\partial P]/(\Theta))_{i-1} \to (\mathbb{R}[\partial P]/(\Theta))_i$ is injective, and hence that $h_{i-1}(P) \leq h_i(P)$ for all $i \leq d/2$, as desired.

Extending Theorem 7.2 to all simplicial spheres is a major open problem in the field: it is a part of the celebrated g-conjecture. The main obstacle is that all known

proofs of the g-theorem for simplicial polytopes, see [49, 50, 74], rely heavily on such tools as toric varieties and the hard Lefschetz theorem or polytopal algebras and the Hodge–Riemann–Minkowski quadratic inequalities between mixed volumes, that is, tools that are (at least at present) available only for convex polytopes.

To prove Theorem 7.3, a few additional ideas are needed. Let P be a cs simplicial d-polytope or, more generally, let (Δ, ϕ) be a cs CM complex over $\mathbb{R}$ with vertex set V. Then the involution ϕ induces the map $\sigma : x_v \to x_{\phi(v)}$ on the set of variables of $S = \mathbb{R}[x_v \ : \ v \in V]$, which, in turn, extends to a unique automorphism σ on S. Furthermore, since $\sigma(I_\Delta) = I_\Delta$, the map σ gives rise to an automorphism on $\mathbb{R}[\Delta]$, which we also denote by σ. The main insight of [75] is that this allows us to equip $\mathbb{R}[\Delta]$ with a finer grading by $\mathbb{N} \times \mathbb{Z}/2\mathbb{Z}$: indeed, define

$$\mathbb{R}[\Delta]_{(i,+1)} := \{f \in \mathbb{R}[\Delta]_i \ : \ \sigma(f) = f\} \quad \text{and} \quad \mathbb{R}[\Delta]_{(i,-1)} := \{f \in \mathbb{R}[\Delta]_i \ : \ \sigma(f) = -f\}.$$

Then $\mathbb{R}[\Delta]_i = \mathbb{R}[\Delta]_{(i,+1)} \oplus \mathbb{R}[\Delta]_{(i,-1)}$ as vector spaces while $\mathbb{R}[\Delta]_{(i,\epsilon_1)} \cdot \mathbb{R}[\Delta]_{(j,\epsilon_2)} \subseteq \mathbb{R}[\Delta]_{(i+j,\epsilon_1\epsilon_2)}$ for all $i, j \in \mathbb{N}$ and $\epsilon_1, \epsilon_2 \in \{\pm 1\}$. It is also not hard to see that $\dim_R \mathbb{R}[\Delta]_{(i,+1)} = \dim_R \mathbb{R}[\Delta]_{(i,-1)} = \frac{1}{2} \dim_R \mathbb{R}[\Delta]_i$ for all $i \geq 1$, and that $\mathbb{R}[\Delta]_{(0,+1)} = \mathbb{R}[\Delta]_0 = \mathbb{R}$.

One can use the Kind–Kleinschmidt criterion [77, Theorem III.2.4] to show that $\mathbb{R}[\Delta]$ has an l.s.o.p. $\theta_1, \ldots, \theta_d$ such that $\theta_j \in \mathbb{R}[\Delta]_{(1,-1)}$ for all j. (For instance, for a cs simplicial polytope P and $\Delta = \partial P$, the l.s.o.p. described in the proof of Theorem 7.2 does the job.) Analyzing the exact sequences as in the proof of Theorem 6.1, see Eq. (6.2), but using our finer grading and, in particular, that $\cdot\theta_j$ maps the degree $(i-1, \pm 1)$ pieces of $\mathbb{R}[\Delta]/(\theta_1, \ldots, \theta_{j-1})$ to the degree $(i, \mp 1)$ pieces, then yields that

$$\dim_{\mathbb{R}} (\mathbb{R}[\Delta]/(\Theta))_{(i,-1)} = \frac{1}{2}\left(h_i(\Delta) - \binom{d}{i}\right) \quad \text{for all } 0 \leq i \leq d; \tag{7.1}$$

see [75] for more details. This establishes the following result of Stanley [75]:

Theorem 7.4. *Let Δ be a cs $\mathbb{R}$-CM complex of dimension $d-1$. Then $h_i(\Delta) \geq \binom{d}{i}$ for all i.*

We now consider the case of $\Delta = \partial P$, where P is a cs simplicial d-polytope (with rational vertices), and $\theta_1, \ldots, \theta_d$ are defined using the coordinates of vertices of P as in the sketch of the proof of Theorem 7.2; in particular, $\theta_j \in \mathbb{R}[\Delta]_{(1,-1)}$. Observe that $\omega = \sum_{v \in V} x_v \in \mathbb{R}[\Delta]_{(1,+1)}$ and hence that $\cdot\omega$ maps $(\mathbb{R}[\Delta]/(\Theta))_{(i-1,-1)}$ to $(\mathbb{R}[\Delta]/(\Theta))_{(i,-1)}$. As the map $\cdot\omega : (\mathbb{R}[\Delta]/(\Theta))_{i-1} \to (\mathbb{R}[\Delta]/(\Theta))_i$ is injective for all $i \leq d/2$, its restriction to $(\mathbb{R}[\Delta]/(\Theta))_{(i-1,-1)}$ is also injective. This, along with Eq. (7.1), completes the proof of Theorem 7.3. It is worth pointing out that Theorem 7.3 was extended by Adin [3, 4] to the classes of all rational simplicial polytopes with a fixed-point-free linear symmetry of prime and prime-power orders. Furthermore, A'Campo-Neuen [1] extended Theorem 7.3 to *toric* g-numbers of *all* cs polytopes (including non-rational non-simplicial ones).

The Murai–Nevo proof [55] of the equality case of Theorem 7.2 involves a beautiful blend of tools such as Alexander duality, Stanley–Reisner rings, the Cohen–Macaulay property, as well as generic initial ideals, and, in particular, Green's crystalization principle [30, Proposition 2.28]. A different proof of both parts of Theorem 7.2, including a sharper version of the inequality part, was very recently obtained by Adiprasito [5, Cor. 6.5 and §7]. His proof relies on Lee's generalized stress spaces [45].

When does equality hold in Theorem 7.3? That is, for a fixed $r \leq d/2$, is there a characterization of cs simplicial d-polytopes with $g_r = \binom{d}{r} - \binom{d}{r-1}$? In a recent paper [41], Klee, Nevo, Novik, and Zheng provide such a characterization in the $r = 2$ case and a conjectural characterization in the $r > 2$ case. Both statements strongly parallel the equality cases of Theorems 7.1 and 7.2. If P is a cs simplicial d-polytope with $d \geq 4$, then we can apply to P the *symmetric stacking operation*: this operation repeatedly attaches (shallow) simplices along antipodal pairs of facets. Note that symmetric stacking preserves both central symmetry and g_2.

Theorem 7.5. *Let P be a cs simplicial d-polytope with $d \geq 4$. Then $g_2(P) = \binom{d}{2} - d$ if and only if P is obtained from $\mathcal{C}_d^*$ by symmetric stacking.*

Conjecture 7.6. *Let P be a cs simplicial d-polytope, and assume that $g_r(P) = \binom{d}{r} - \binom{d}{r-1}$ for some $3 \leq r \leq \lfloor d/2 \rfloor$. Then there exists a unique polytopal complex $\mathcal{C}$ in $\mathbb{R}^d$ with the following properties: (i) one of the faces of $\mathcal{C}$ is the cross-polytope $\mathcal{C}_d^*$, all other faces of $\mathcal{C}$ are simplices that come in antipodal pairs, (ii) $\mathcal{C}$ is a "cellulation" of P, that is, $\bigcup_{C \in \mathcal{C}} C = P$, and (iii) each element $C \in \mathcal{C}$ of dimension $\leq d - r$ is a face of P.*

We mention that [41, Conjecture 8.6] provides a more detailed version of the above conjecture, and we refer our readers to Ziegler's book [84, Section 8.1] for the definition of a *polytopal* complex. Conjecture 7.6, if true, would imply the following weaker statement that is also wide open at present. (The $r = 2$ case does hold; this is immediate from Theorem 7.5.)

Conjecture 7.7. *Let P be a cs simplicial d-polytope. If $g_r(P) = g_r(\mathcal{C}_d^*)$ for some $3 \leq r \leq \lfloor d/2 \rfloor - 1$, then $g_{r+1}(P) = g_{r+1}(\mathcal{C}_d^*)$.*

One consequence of Theorems 7.3 and 7.5, along with the fact that the f-numbers of spheres are nonnegative linear combinations of the g-numbers, is that among all cs simplicial d-polytopes with n vertices, a polytope obtained from $\mathcal{C}_d^*$ by symmetric stacking simultaneously minimizes all the face numbers; furthermore, if $d \geq 4$, then such polytopes are the only minimizers.

While we will not discuss here the proof of Theorem 7.5, it is worth pointing out that it relies on the rigidity theory of frameworks (some ingredients of this theory are briefly outlined in the next section), and, in particular, on (1) the theorem of Whiteley [82] asserting that the graph of a simplicial d-polytope ($d \geq 3$) with its natural embedding in $\mathbb{R}^d$ is infinitesimally rigid, and on (2) the fact that if P is a cs simplicial d-polytope with $g_2(P) = \binom{d}{2} - d$, then *all* stresses on P must be symmetric (i.e., assign the same weight to each edge and its antipode.) This latter

observation follows from work of Stanley [75] and Lee [45], and also from more recent work of Sanyal, Werner, and Ziegler [68, Theorem 2.1].

To summarize: in contrast with the upper bound-type results, an analog of the GLBT for cs simplicial polytopes (at least its inequality part developed in Theorem 7.3) is a very well-understood part of the story. Furthermore, in the case of the LBT we even have a characterization of the minimizers (see Theorem 7.5). The part that is still missing is a characterization of cs simplicial d-polytopes with $g_r = \binom{d}{r} - \binom{d}{r-1}$ for $3 \leq r \leq \lfloor d/2 \rfloor$. Conjecture 7.6 proposes such a characterization.

8 The Lower Bound Conjecture for cs Spheres and Manifolds

The final part of our story concerns a conjectural analog of the LBT for cs spheres (manifolds and even normal pseudomanifolds)—a necessary step in our quest for a cs analog of the g-conjecture. To this end, it is worth recalling that in the world of simplicial complexes without a symmetry assumption, where at present the GLBT (Theorem 7.2) is only known to hold for the class of simplicial polytopes, works of Walkup [80], Barnette [9], Kalai [37], Fogelsanger [24], and Tay [78] show that the LBT (Theorem 7.1) holds much more generally:

Theorem 8.1. *Let Δ be a simplicial complex of dimension $d-1 \geq 3$. Assume further that Δ is a connected simplicial manifold (or even a normal pseudomanifold). Then $g_2(\Delta) \geq 0$, with equality if and only if Δ is the boundary complex of a stacked polytope.*

In view of this result, we strongly suspect that Stanley's inequality on the g_2-number of cs polytopes (see Theorem 7.3) as well as the characterization of the minimizers given in Theorem 7.5 continue to hold in the generality of cs simplicial spheres or perhaps even cs normal pseudomanifolds. This conjecture is however wide open at present.

Conjecture 8.2. *Let Δ be a cs simplicial complex of dimension $d-1 \geq 3$. Assume further that Δ is a simplicial sphere (or a connected simplicial manifold or even a normal pseudomanifold). Then $g_2(\Delta) \geq \binom{d}{2} - d$. Furthermore, equality holds if and only if Δ is the boundary complex of a cs d-polytope obtained from the cross-polytope $\mathcal{C}_d^*$ by symmetric stacking.*

In the rest of this short section, we discuss one potential approach to attacking this conjecture: via the rigidity theory of frameworks. The (now wide) use of this theory in the study of face numbers of simplicial complexes was pioneered by Kalai in his celebrated proof of Theorem 8.1 for simplicial manifolds [37]. Below we review several results and definitions pertaining to this fascinating theory. We refer our readers to Asimow and Roth [6, 7] for a friendly introduction to this subject.

Let $G = (V, E)$ be a finite graph. A map $\rho : V \to \mathbb{R}^d$ is called a *d-embedding* of G if $\text{aff}\{\rho(v) \ : \ v \in V\} = \mathbb{R}^d$. The graph G, together with a d-embedding ρ, is called a *d-framework*. An infinitesimal motion of a d-framework (G, ρ) is a map $\mu : V(G) \to \mathbb{R}^d$ such that for any edge $\{u, v\}$ in G,

$$\frac{d}{dt}\bigg|_{t=0} \big\| (\rho(u) + t\mu(u)) - (\rho(v) + t\mu(v)) \big\|^2 = 0.$$

An infinitesimal motion μ of (G, ρ) is trivial if $\frac{d}{dt}\Big|_{t=0} \big\| (\rho(u) + t\mu(u)) - (\rho(v) + t\mu(v)) \big\|^2 = 0$ holds for every two vertices u, v of G. (Trivial infinitesimal motions correspond to a start of an isometric motion of $\mathbb{R}^d$.) We say that a d-framework (G, ρ) is *infinitesimally rigid* if every infinitesimal motion μ of (G, ρ) is trivial.

A *stress* on a d-framework (G, ρ) is an assignment of weights $\omega = (\omega_e \ : \ e \in E(G))$ to the edges of G such that for each vertex v equilibrium holds:

$$\sum_{u \ : \ \{u,v\} \in E(G)} \omega_{\{u,v\}}(\rho(v) - \rho(u)) = \mathbf{0}.$$

We denote the space of all stresses on (G, ρ) by $\mathcal{S}(G, \rho)$. The stresses on (G, ρ) correspond to the elements in the kernel of the transpose of a certain $f_1(G) \times df_0(G)$ matrix $\text{Rig}(G, \rho)$ known as the *rigidity matrix* of (G, ρ).

The relevance of rigidity theory to the Lower Bound Theorem is explained by the following fundamental fact that is an easy consequence of the Implicit Function Theorem (see [6, 7]).

Theorem 8.3. *Let (G, ρ) be a d-framework. Then the following statements are equivalent:*

- (G, ρ) *is infinitesimally rigid;*
- rank $\text{Rig}(G, \rho) = df_0(G) - \binom{d+1}{2}$*;*
- $\dim_{\mathbb{R}} \mathcal{S}(G, \rho) = f_1(G) - df_0(G) + \binom{d+1}{2}$.

A graph G is called *generically d-rigid* if there exists a d-embedding ρ of G such that (G, ρ) is infinitesimally rigid; in this case, the set of infinitesimally rigid d-embeddings of G is an open dense subset of the set of all d-embeddings. Recall that if Δ is a $(d-1)$-dimensional simplicial complex, then $g_2(\Delta) = h_2(\Delta) - h_1(\Delta) = f_1(\Delta) - df_0(\Delta) + \binom{d+1}{2}$. The last condition of Theorem 8.3 then implies that if the graph (i.e., 1-skeleton) of Δ is generically d-rigid, then $g_2(\Delta) \geq 0$.

Two basic but very useful results in rigidity theory are the gluing lemma ([7, Theorem 2] and [83, Lemma 11.1.9]) and the cone lemma [81]. The gluing lemma asserts that if two graphs G_1 and G_2 are generically d-rigid and share at least d vertices, then their union $G_1 \cup G_2$ is also generically d-rigid, while the cone lemma posits that G is generically d-rigid if and only if the graph of the cone over G is generically $(d+1)$-rigid. Additionally, by Gluck's result [29], the graph of any simplicial 2-sphere is generically 3-rigid. Now, if Δ is a simplicial $(d-1)$-manifold

(with $d \geq 4$) and F is a $(d-4)$-face of Δ, then the link of F in Δ is a simplicial 2-sphere and hence has a generically 3-rigid graph. This observation, along with the gluing and cone lemmas, allowed Kalai [37] to prove by induction on $d \geq 4$ that the graph of any simplicial $(d-1)$-manifold is generically d-rigid and thus to establish the inequality part of Theorem 8.1 for all simplicial manifolds.

The relationship of infinitesimal rigidity to the Stanley–Reisner ring was worked out in [45]. Specifically, let Δ be a $(d-1)$-dimensional simplicial complex with vertex set V and let $\rho : V \to \mathbb{R}^d$ be a d-embedding of the graph of Δ, $G(\Delta)$. Define d linear forms in $\mathbb{R}[x_v \ : \ v \in V]$ by $\theta_j := \sum_{v \in V} \rho(v)_j x_v$ for $j = 1, \dots, d$, where $\rho(v)_j$ denotes the jth coordinate of $\rho(v) \in \mathbb{R}^d$, and let $\theta_{d+1} := \sum_{v \in V} x_v$ (cf. Sect. 7, and especially the sketch of the proof of Theorem 7.2). Theorem 10 of [45] implies that if $(G(\Delta), \rho)$ is infinitesimally rigid, then $\dim_{\mathbb{R}}(\mathbb{R}[\Delta]/(\theta_1, \dots, \theta_d, \theta_{d+1}))_2 = \dim_{\mathbb{R}} \mathcal{S}(G(\Delta), \rho) = g_2(\Delta)$. Equivalently, if $(G(\Delta), \rho)$ is infinitesimally rigid, then $\theta_{i+1} : (\mathbb{R}[\Delta]/(\theta_1, \dots, \theta_i))_1 \to (\mathbb{R}[\Delta]/(\theta_1, \dots, \theta_i))_2$ is an injection for all $0 \leq i \leq d$.

Assume now that Δ is centrally symmetric. If there is a d-embedding $\rho : V(\Delta) \to \mathbb{R}^d$ that respects symmetry and such that $(G(\Delta), \rho)$ is infinitesimally rigid, then the previous paragraph along with the $\mathbb{N} \times \mathbb{Z}/2\mathbb{Z}$-grading of $\mathbb{R}[\Delta]$ and relevant computations of Sect. 7 (e.g., Eq. (7.1)) shows that $g_2(\Delta) \geq \binom{d}{2} - d$. (This also follows from the argument in [68, Section 2].) In particular, the following conjecture, if true, would imply the inequality part of Conjecture 8.2.

Conjecture 8.4. *Let Δ be a $(d-1)$-dimensional cs simplicial complex with an involution ϕ. Assume further that Δ is a simplicial sphere (or a connected simplicial manifold or even a normal pseudomanifold) and that $d-1 \geq 3$. Then there exists a d-embedding $\rho : V(\Delta) \to \mathbb{R}^d$ that respects symmetry, i.e., $\rho(\phi(v)) = -\rho(v)$ for each vertex v of Δ, and such that $(G(\Delta), \rho)$ is infinitesimally rigid.*

For instance, if Δ is the boundary complex of a cs simplicial d-polytope $P \subset \mathbb{R}^d$, then by Whiteley's theorem [82], the natural d-embedding of the graph of P qualifies for ϕ as in Conjecture 8.4.

One of the reasons Conjecture 8.4 appears to be hard in the general case is that the links of cs complexes are usually not centrally symmetric, and so starting with *cs* 2-spheres as the base case (and proceeding by induction via the cone and gluing lemmas) does not work. Conjecture 8.4 of [41] proposes a statement about rigidity of graphs of (non-cs) simplicial 2-spheres that, if true, will provide an appropriate base case, and imply Conjecture 8.4. In any case, as a step in our quest for a cs analog of the g-conjecture, it would be extremely interesting to shed any light on Conjectures 8.2 and 8.4 as well as to attempt to strengthen the inequality of Conjecture 8.2 in the spirit of results from [54, Theorem 5.3(i)] and [56]. Such strengthened inequalities would provide lower bounds on g_2 of a cs simplicial manifold (or even a normal pseudomanifold) Δ in terms of the first homology or perhaps even in terms of the fundamental group of Δ and/or of the $\mathbb{Z}/2\mathbb{Z}$-quotient of Δ.

9 Concluding Remarks

The recent decades made the theory of face numbers into a very active and a rather large field. Consequently, there are quite a few topics we glossed over or omitted in this paper. Among them are the face numbers of general (not necessarily simplicial) cs polytopes. However, one of the conjectures about general cs polytopes we cannot avoid mentioning is Kalai's "3^d-conjecture" [39, Conjecture A]: it posits that the total number of faces in a cs d-polytope (including the empty face but excluding the polytope itself) is at least 3^d. If the conjecture is true, there are multiple minimizers: the class of polytopes with exactly 3^d faces includes at least all the Hanner polytopes. (If Mahler's conjecture holds, then Hanner polytopes are the minimizers of Mahler volume among all the cs convex bodies.) At present the 3^d-conjecture is known to hold for all cs simplicial polytopes (this is an immediate corollary of Theorem 7.4) and, by duality, all simple polytopes, as well as for all at most 4-dimensional cs polytopes [68]. The conjecture is wide open in all other cases.

We have also only barely touched on the face numbers of cs simplicial manifolds and instead concentrated on the face numbers of cs simplicial polytopes and spheres. Papers [59], [61, Section 4], and [42] contain some results pertaining to the Upper Bound Theorem for cs simplicial manifolds as well as to Sparla's conjecture [71, Conjecture 4.12], [72] on the Euler characteristic of even-dimensional cs simplicial manifolds.

While we had to skip quite a few of the topics and could not possibly do justice to all of the existing methods, we hope we have conveyed at least some of the essence and beauty of this fascinating subject! We are very much looking forward to progress on the many existing as well as yet unstated problems about cs polytopes and simplicial complexes, and to new interactions between combinatorics, discrete geometry, commutative algebra, and geometric analysis that will lead to this progress!

Acknowledgements I am grateful to Steve Klee, Connor Sawaske, Hailun Zheng, Günter Ziegler, and the anonymous referee for numerous comments on the preliminary version of this paper.

References

1. A. A'Campo-Neuen, On toric h-vectors of centrally symmetric polytopes. Arch. Math. (Basel) **87**(3), 217–226 (2006)
2. R.M. Adin, Combinatorial structure of simplicial complexes with symmetry. Ph.D. thesis, Hebrew University, Jerusalem (1991)
3. R.M. Adin, On h-vectors and symmetry, *Jerusalem Combinatorics '93*, vol. 178, Contemporary in Mathematics (American Mathematical Society, Providence, 1994), pp. 1–20
4. R.M. Adin, On face numbers of rational simplicial polytopes with symmetry. Adv. Math. **115**(2), 269–285 (1995)
5. K. Adiprasito, Toric chordality. J. Math. Pures Appl. (9) **108**(5), 783–807 (2017)
6. L. Asimow, B. Roth, The rigidity of graphs. Trans. Am. Math. Soc. **245**, 279–289 (1978)
7. L. Asimow, B. Roth, The rigidity of graphs II. J. Math. Anal. Appl. **68**(1), 171–190 (1979)

8. K. Ball, An elementary introduction to modern convex geometry, in *Flavors of Geometry*, vol. 31, Mathematical Sciences Research Institute Publications (Cambridge University Press, Cambridge, 1997)
9. D. Barnette, Graph theorems for manifolds. Isr. J. Math. **16**, 62–72 (1973)
10. D. Barnette, A proof of the lower bound conjecture for convex polytopes. Pac. J. Math. **46**, 349–354 (1973)
11. A. Barvinok, *A Course in Convexity*, vol. 54, Graduate Studies in Mathematics (American Mathematical Society, Providence, 2002)
12. A. Barvinok, S.J. Lee, I. Novik, Explicit constructions of centrally symmetric k-neighborly polytopes and large strictly antipodal sets. Discret. Comput. Geom. **49**(3), 429–443 (2013)
13. A. Barvinok, S.J. Lee, I. Novik, Neighborliness of the symmetric moment curve. Mathematika **59**(1), 223–249 (2013)
14. A. Barvinok, I. Novik, A centrally symmetric version of the cyclic polytope. Discret. Comput. Geom. **39**(1–3), 76–99 (2008)
15. L.J. Billera, C.W. Lee, A proof of the sufficiency of McMullen's conditions for f-vectors of simplicial convex polytopes. J. Comb. Theory Ser. A **31**, 237–255 (1981)
16. A. Björner, A. Paffenholz, J. Sjöstrand, G.M. Ziegler, Bier spheres and posets. Discret. Comput. Geom. **34**(1), 71–86 (2005)
17. J. Browder, I. Novik, Face numbers of generalized balanced Cohen-Macaulay complexes. Combinatorica **31**(6), 669–689 (2011)
18. G.R. Burton, The nonneighbourliness of centrally symmetric convex polytopes having many vertices. J. Comb. Theory Ser. A **58**, 321–322 (1991)
19. C. Carathéodory, Über den Variabilitatsbereich det Fourierschen Konstanten von positiven harmonischen Furktionen. Rend. Circ. Mat. Palermo **32**, 193–217 (1911)
20. V.I. Danilov, The geometry of toric varieties. Uspekhi Mat. Nauk **33**(2(200)), 85–134, 274 (1978)
21. L. Danzer, B. Grünbaum, Über zwei probleme bezüglich konvexer körper von P. Erdös und von V. L. Klee. Math. Z. **79**, 95–99 (1962)
22. D.L. Donoho, High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. Discret. Comput. Geom. **35**(4), 617–652 (2006)
23. D.L. Donoho, J. Tanner, Exponential bounds implying construction of compressed sensing matrices, error-correcting codes, and neighborly polytopes by random sampling. IEEE Trans. Inf. Theory **56**(4), 2002–2016 (2010)
24. A. Fogelsanger, The generic rigidity of minimal cycles. Ph.D. thesis, Cornell University (1988)
25. G. Freiman, E. Lipkin, L. Levitin, A polynomial algorithm for constructing families of k-independent sets. Discret. Math. **70**(2), 137–147 (1988)
26. W. Fulton, Introduction to toric varieties, in *The William H. Roever Lectures in Geometry*, vol. 131, Annals of Mathematics Studies (Princeton University Press, Princeton, 1993)
27. D. Gale, Neighborly and cyclic polytopes, in *Proceedings of the Seventh Symposium in Pure Mathematics, Vol. VII*. (American Mathematical Society, Providence, 1963), pp. 225–232
28. A.Y. Garnaev, E.D. Gluskin, The widths of a Euclidean ball. Dokl. Akad. Nauk SSSR **277**(5), 1048–1052 (1984). (English translation: Sov. Math. Dokl. **30**(1), 200–204 (1984))
29. H. Gluck, Almost all simply connected closed surfaces are rigid, in *Geometric Topology (Proceedings of the Geometric Topology Conference, Park City, Utah, 1974)*, Lecture Notes in Mathematics, vol. 438 (Springer, Berlin, 1975), pp. 225–239
30. M.L. Green, Generic initial ideals, in *Six Lectures on Commutative Algebra (Bellaterra, 1996)*, vol. 166, Progress in Mathematics (Birkhäuser, Basel, 1998), pp. 119–186
31. B. Grünbaum, *Convex Polytopes*, vol. 221, 2nd edn., Graduate Texts in Mathematics (Springer, New York, 2003). Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M, Ziegler
32. B. Grünbaum, T.S. Motzkin, On polyhedral graphs, in *Proceedings of Symposia in Pure Mathematics, Vol. VII* (American Mathematical Society, Providence, 1963), pp. 285–290
33. E.R. Halsey, Zonotopal complexes on the d-cube. Ph.D. thesis, University of Washington (1972)

34. P. Hersh, I. Novik, A short simplicial h-vector and the upper bound theorem. Discret. Comput. Geom. **28**(3), 283–289 (2002)
35. M. Hochster, Cohen-Macaulay rings, combinatorics, and simplicial complexes, in *Ring Theory, II (Proceedings of the Second Conference, Univ. Oklahoma, Norman, Okla., 1975)*, vol. 26, Lecture Notes in Pure and Applied Mathematics (Dekker, New York, 1977), pp. 171–223
36. W. Jockusch, An infinite family of nearly neighborly centrally symmetric 3-spheres. J. Comb. Theory Ser. A **72**(2), 318–321 (1995)
37. G. Kalai, Rigidity and the lower bound theorem I. Invent. Math. **88**, 125–151 (1987)
38. G. Kalai, Many triangulated spheres. Discret. Comput. Geom. **3**(1), 1–14 (1988)
39. G. Kalai, The number of faces of centrally-symmetric polytopes. Graphs Comb. **5**(1), 389–391 (1989)
40. B.S. Kašin, The widths of certain finite-dimensional sets and classes of smooth functions. Izv. Akad. Nauk SSSR Ser. Mat. **41(2): 334–351**, 478 (1977)
41. S. Klee, E. Nevo, I. Novik, H. Zheng, A lower bound theorem for centrally symmetric simplicial polytopes. Discret. Comput. Geom. (2018). https://doi.org/10.1007/s00454-018-9978-z
42. S. Klee, I. Novik, Centrally symmetric manifolds with few vertices. Adv. Math. **229**, 487–500 (2012)
43. V. Klee, A combinatorial analogue of Poincaré's duality theorem. Can. J. Math. **16**, 517–531 (1964)
44. V. Klee, On the number of vertices of a convex polytope. Can. J. Math. **16**, 701–720 (1964)
45. C.W. Lee, Generalized stress and motions, in *Polytopes: Abstract, Convex And Computational (Scarborough, ON, 1993)*, vol. 440, NATO Advanced Study Series C Institute Mathematical and Physical Sciences (Kluwer Academic Publishers, Dordrecht, 1994), pp. 249–271
46. N. Linial, I. Novik, How neighborly can a centrally symmetric polytope be? Discret. Comput. Geom. **36**, 273–281 (2006)
47. F. Lutz, Triangulated manifolds with few vertices and vertex-transitive group actions. Dissertation, Technischen Universität Berlin, Berlin (1999)
48. P. McMullen, The maximum numbers of faces of a convex polytope. Mathematika **17**, 179–184 (1970)
49. P. McMullen, On simple polytopes. Invent. Math. **113**(2), 419–444 (1993)
50. P. McMullen, Weights on polytopes. Discret. Comput. Geom. **15**(4), 363–388 (1996)
51. P. McMullen, G.C. Shephard, Diagrams for centrally symmetric polytopes. Mathematika **15**, 123–138 (1968)
52. P. McMullen, D.W. Walkup, A generalized lower-bound conjecture for simplicial polytopes. Mathematika **18**, 264–273 (1971)
53. T.S. Motzkin, Comonotone curves and polyhedra. Bull. Am. Math. Soc. **63**, 35 (1957)
54. S. Murai, Tight combinatorial manifolds and graded Betti numbers. Collect. Math. **66**(3), 367–386 (2015)
55. S. Murai, E. Nevo, On the generalized lower bound conjecture for polytopes and spheres. Acta Math. **210**(1), 185–202 (2013)
56. S. Murai, I. Novik, Face numbers and the fundamental group. Isr. J. Math. **222**(1), 297–315 (2017)
57. E. Nevo, F. Santos, S. Wilson, Many triangulated odd-dimensional spheres. Math. Ann. **364**(3–4), 737–762 (2016)
58. I. Novik, Upper bound theorems for homology manifolds. Isr. J. Math. **108**, 45–82 (1998)
59. I. Novik, On face numbers of manifolds with symmetry. Adv. Math. **192**, 183–208 (2005)
60. I. Novik, From acute sets to centrally symmetric 2-neighborly polytopes. SIAM. J. Discret. Math. **32**, 1572–1576 (2018)
61. I. Novik, E. Swartz, Socles of buchsbaum modules, posets and complexes. Adv. Math. **222**, 2059–2084 (2009)
62. I. Novik, E. Swartz, Face numbers of pseudomanifolds with isolated singularities. Math. Scand. **110**(2), 198–222 (2012)
63. A. Padrol, Many neighborly polytopes and oriented matroids. Discret. Comput. Geom. **50**(4), 865–902 (2013)

64. J. Pfeifle, Extremal constructions for polytopes and spheres. Dissertation, Technischen Universität Berlin, Berlin (2003)
65. J. Pfeifle, G.M. Ziegler, Many triangulated 3-spheres. Math. Ann. **330**(4), 829–837 (2004)
66. G.A. Reisner, Cohen-Macaulay quotients of polynomial rings. Adv. Math. **21**(1), 30–49 (1976)
67. M. Rudelson, R. Vershynin, Geometric approach to error correcting codes and reconstruction of signals. Int. Math. Res. Not. **64**, 4019–4041 (2005)
68. R. Sanyal, A. Werner, G.M. Ziegler, On Kalai's conjectures concerning centrally symmetric polytopes. Discret. Comput. Geom. **41**(2), 183–198 (2009)
69. R. Schneider, Neighbourliness of centrally symmetric polytopes in high dimensions. Mathematika **22**, 176–181 (1975)
70. Z. Smilansky, Convex hulls of generalized moment curves. Isr. J. Math. **52**(1–2), 115–128 (1985)
71. E. Sparla, Geometrische und kombinatorische Eigenschaften triangulierter Mannigfaltigkeiten. Berichte aus der Mathematik. [Reports from Mathematics]. Verlag Shaker, Aachen. Dissertation, Universität Stuttgart, Stuttgart (1997)
72. E. Sparla, An upper and a lower bound theorem for combinatorial 4-manifolds. Discret. Comput. Geom. **19**, 575–593 (1998)
73. R.P. Stanley, The upper bound conjecture and Cohen-Macaulay rings. Stud. Appl. Math. **54**, 135–142 (1975)
74. R.P. Stanley, The number of faces of a simplicial convex polytope. Adv. Math. **35**, 236–238 (1980)
75. R.P. Stanley, On the number of faces of centrally-symmetric simplicial polytopes. Graphs Comb. **3**, 55–66 (1987)
76. R.P. Stanley, A monotonicity property of h-vectors and h^*-vectors. Eur. J. Comb. **14**(3), 251–258 (1993)
77. R.P. Stanley, *Combinatorics and Commutative Algebra*, 2nd edn., Progress in Mathematics (Birkhäuser, Boston, 1996)
78. T.-S. Tay, Lower-bound theorems for pseudomanifolds. Discret. Comput. Geom. **13**(2), 203–216 (1995)
79. C. Vinzant, Edges of the Barvinok-Novik orbitope. Discret. Comput. Geom. **46**(3), 479–487 (2011)
80. D.W. Walkup, The lower bound conjecture for 3- and 4-manifolds. Acta Math. **125**, 75–107 (1970)
81. W. Whiteley, Cones, infinity and 1-story buildings. Struct. Topol. **8**, 53–70 (1983)
82. W. Whiteley, Infinitesimally rigid polyhedra. I. Statics of frameworks. Trans. Am. Math. Soc. **285**(2), 431–465 (1984)
83. W. Whiteley, Some matroids from discrete applied geometry, in *Matroid Theory (Seattle, WA, 1995)*, vol. 197, Contemporary in Mathematics (American Mathematical Society, Providence, 1996), pp. 171–311
84. G.M. Ziegler, *Lectures on Polytopes*, vol. 152, Graduate Texts in Mathematics (Springer, New York, 1995)

Crystal Constructions in Number Theory

Anna Puskás

Abstract Weyl group multiple Dirichlet series and metaplectic Whittaker functions can be described in terms of crystal graphs. We present crystals as parameterized by Littelmann patterns, and we give a survey of purely combinatorial constructions of prime power coefficients of Weyl group multiple Dirichlet series and metaplectic Whittaker functions using the language of crystal graphs. We explore how the branching structure of crystals manifests in these constructions, and how it allows access to some intricate objects in number theory and related open questions using tools of algebraic combinatorics.

1 Introduction

Crystal graphs are combinatorial objects appearing in the representation theory of semisimple Lie algebras. To an irreducible representation of a semisimple Lie algebra $\mathfrak{g}$, one may associate a crystal graph $\mathcal{C}$. The vertices of this graph are in bijection with a weight basis of the representation, and the edges are colored by a set of simple roots of $\mathfrak{g}$.

Crystals were first studied in connection with the representation theory of the quantized universal enveloping algebra. However, in this chapter it is their structure as a colored (directed) graph and their symmetries related to the Weyl group of $\mathfrak{g}$ that are of interest to us. Crystals turn out to be a valuable tool in constructing certain objects from number theory: coefficients of multiple Dirichlet series and metaplectic Whittaker functions.

Interest in multiple Dirichlet series and metaplectic Whittaker functions is motivated by hard questions in analytic number theory, for example, the Lindelöf Hypothesis, and the study of automorphic forms [20]. The relevant literature in number theory is extensive (see Sect. 1.2). However, since these objects have constructions that are

A. Puskás (✉)
Kavli Institute for the Physics and Mathematics of the Universe, Kashiwa, Chiba, Japan
e-mail: anna.puskas@ipmu.jp

H. Barcelo et al. (eds.), *Recent Trends in Algebraic Combinatorics*,
Association for Women in Mathematics Series 16,
https://doi.org/10.1007/978-3-030-05141-9_10

almost purely combinatorial in nature, their study can be approached using tools of algebraic combinatorics.

In this chapter, we are interested in this approach. Our main goal is to present combinatorial constructions of metaplectic Whittaker functions and coefficients of multiple Dirichlet series corresponding to root systems of the four infinite families of Cartan types. To do so, we use the language of Littelmann patterns. We highlight how the branching structure of crystals is apparent in the constructions and indicate how this aspect turns out to be significant in the study of the related questions from number theory.

Before giving an overview of the structure of the chapter, we say a few more words on the relevant objects.

Crystal graphs can be parameterized (without referring to the representation theory of the quantum group) using a variety of combinatorial devices, such as the Littelmann path model, Gelfand–Tsetlin patterns, Lusztig's parametrization [35, 36], tableaux of Lakshmibai and Sheshadri [32] or Kashiwara and Nakashima [31]. For a thorough introduction to the theory of crystals from a combinatorial perspective, the reader is encouraged to consult [18].

Here we present crystals in terms of Berenstein–Zelevinsky–Littelmann paths and Littelmann patterns [34]. Our reason for this choice is twofold. First, most of the constructions in number theory that we are concerned with were either originally given in this language, or are easily rephrased in such terms. Moreover, phrasing the constructions in terms of Littelmann patterns highlights the role of the branching structure of crystals (as well as the significance of some "nice elements" of the Weyl group) very well.

A major hurdle any expository writing on this topic has to overcome is the inherent intricacy and volume of the theory of multiple Dirichlet series and metaplectic Whittaker functions. Since we wish to take a purely combinatorial approach, we largely try and circumvent this issue. Some background on multiple Dirichlet series and Whittaker functions (as well as on metaplectic groups) will be given in Sect. 2.4. For now we say that through their connection to an algebraic group over a local or global field, these objects from number theory are related to the representation theory of the underlying Lie algebra $\mathfrak{g}$. Their constructions involve producing, for a dominant weight λ, a polynomial $P_\lambda(\mathbf{x})$ in r variables, where r is the rank of the Lie algebra $\mathfrak{g}$. In Sect. 2.4, we shall say more about how a Weyl group multiple Dirichlet series or a metaplectic Whittaker function gives rise to such a polynomial P_λ. However, for most of the chapter we shall ignore details of this background and concern ourselves with producing a polynomial P_λ as a sum over a crystal graph. The "constructions" mentioned throughout the chapter refer to constructions of a polynomial P_λ, depending on the context, this may agree with a p-part of a Weyl group multiple Dirichlet series or certain values of a spherical Whittaker function.

The combinatorial perspective of focusing our attention on the polynomials P_λ is, on the one hand, helpful when considering questions motivated by the analytic background. On the other hand, these objects are interesting in their own right. This is due to the fact that they can be thought of as deformations of highest weight characters. As a result, techniques of character theory come into play. As a method

to study these polynomials, it provides insight into the original analytical objects. Furthermore, it motivates further questions.

To provide an example, we mention two aspects of the polynomials P_λ now. One is Weyl symmetry: the polynomials P_λ inherit certain functional equations under the Weyl group corresponding to the underlying root system. Hence one may construct such a polynomial by (1) taking a sum over an object that is similarly symmetric, such as a highest weight crystal or (2) by taking an "average" over the Weyl group. (See Sect. 1.2 for relevant results in the literature.) Understanding the relationship between these two approaches is a large part of the motivation between studying the constructions combinatorially, and in the cases where the question is resolved, the branching structure of crystals turns out to play a significant role.

We briefly explain the relevance of character theory. In the simplest special case, P_λ looks very similar to the deformation of a Schur polynomial; more generally, to the deformation of a Weyl character. (For P_λ a Whittaker function, this is a consequence of the Casselman–Shalika formula.) Hence one expects that the behavior of families of polynomials P_λ will be similar to the behavior of characters. On the one hand, this means that identities of Weyl characters may provide a useful tool of study. These come in a couple of different flavors. For example, the Weyl(-Kac) character formula produces a character. Branching rules describe the behavior of characters under restriction. Indeed, Tokuyama's theorem (a deformation of the Weyl character formula) turns out to be key in investigating the relationship of the two approaches (1) and (2) mentioned above. Generalizing it to the polynomials P_λ requires understanding the branching properties of the P_λ. We shall elaborate on this point in Sects. 1.3 and 1.4 below.

On the other hand, one may ask if products of the P_λ satisfy some "enhanced" version of other character identities. For example, does a product of such polynomials satisfy "deformed" Pieri and Littelwood–Richardson rules? Question of this flavor may be investigated using any description of these objects. We shall see that the P_λ can be defined in terms of the combinatorial structure of a crystal and a few Gauss sums. Hence any question about them can be phrased in terms of the crystal structure and identities of Gauss sums.

We give an overview of the structure of the chapter.

1.1 Structure of the Paper

In the remainder of this Introduction, we first give a brief review of results constructing Weyl group multiple Dirichlet series or metaplectic Whittaker functions (Sect. 1.2). In Sect. 1.3, we explain how a theorem of Tokuyama is related to this topic, and how Demazure–Lusztig operators can be used to study, and extend the constructions discussed in this chapter to greater generality. Section 1.4 provides some further insight into the meaning and significance of branching.

Littelmann patterns and their bijection with crystal elements are discussed in Sect. 3. The constructions of Whittaker functions and prime power coefficients of

multiple Dirichlet series in terms of highest weight crystals are presented in Sect. 4, and the relationship of the constructions with the branching structure of crystals is highlighted in Sect. 5.

Section 2 serves to present some preliminaries. We introduce notation (Sect. 2.1) and present Gauss sums, a necessary arithmetic ingredient to the constructions (Sect. 2.2). We then give a brief introduction to crystals and Berenstein–Zelevinsky–Littelmann paths (Sect. 2.3). We also provide a little more insight into how coefficients of multiple Dirichlet series and Weyl group metaplectic Whittaker functions give rise to polynomials $P_\lambda(\mathbf{x})$, related to sums over highest weight crystals (Sect. 2.4).

1.2 A Review of Literature

We discuss the literature of constructions of multiple Dirichlet series and Whittaker functions. Our interest here is from the perspective of combinatorics. Hence we shall focus on the role of the branching structure of the crystals and the significance of special words in the Weyl group. For an insightful and thorough introduction to the topic from a number-theoretic perspective, the reader is encouraged to consult [17], the Introduction of the volume where many of the constructions discussed in this chapter were published. The role of this section is to provide this topic with a wider context; strictly speaking, it is not necessary for the understanding of any of the later parts.

Brubaker, Bump, Chinta, Friedberg, and Hoffstein [12] introduced Weyl group multiple Dirichlet series (WMDS), series in several complex variables with functional equations governed by a finite Weyl group, corresponding to a root system Φ of finite type. As mentioned above, there are two separate approaches to how to associate a WMDS to a root system: by taking a sum over a crystal, or by Chinta and Gunnells [21], by averaging over the Weyl group.

The authors of [12] conjecturally related WMDS to Whittaker coefficients of metaplectic Eisenstein series. This connection is of interest in that it allows one to prove functional equations and analytic continuation of the constructed series. Elucidating this connection motivates study of these objects as well. In the following paragraphs when we refer to a "conjectural description" of a WMDS as a sum over a crystal, we mean either that the constructed series is conjectured to be the Whittaker coefficient of a metaplectic Eisenstein series, or that it is conjectured to agree with a series constructed via the averaging method.

We shall mention relevant results in all four infinite families of Cartan types; some of these results will be covered in more detail in Sect. 4.

Brubaker, Bump, and Friedberg [15] describe the Fourier–Whittaker coefficients of Eisenstein series on a metaplectic cover of the general linear group as a Weyl group multiple Dirichlet series. They compute the prime power coefficients (p-parts) of these series in terms of the string parametrization of a crystal by Berenstein and Zelevinsky [4, 5] and Littelmann [34]. In [14], the same authors further explore the

combinatorics in the type A case. They give two separate constructions of the p-part in type A. These can be seen as corresponding to two different choices of *nice decompositions* of the long element of the Weyl group. The authors then prove that the two descriptions give the same p-parts through a subtle combinatorial argument. The equivalence of the two statements allows them to prove analytic continuation and functional equations for the emerging multiple Dirichlet series. In proving the equivalence, they observe the significance of some purely combinatorial phenomena—such as the Schützenberger involution. Their method provides an example of how to build p-parts of multiple Dirichlet series out of finite crystal data.

Beineke, Brubaker, and Frechette [1, 2] give a definition for a WMDS in terms of statistics on a highest weight crystal of Cartan type C. They prove analytic continuation and functional equation of such series using a connection to Eisenstein series over odd orthogonal groups in the nonmetaplectic case, and conjecture a similar connection in general.

Friedberg and Zhang [26] study Eisenstein series over metaplectic covers of odd orthogonal groups. They then describe the p-parts of the MDS that are the Whittaker coefficients of these series in terms of type C highest weight crystals. They in fact give two descriptions. The first one is only valid in the case of odd covers; this proves the conjectured connection in [2] above. The second is uniform in the degree n of the metaplectic cover, but the assignment of number-theoretic data to the combinatorial structure is more subtle. An interesting feature of their methods is that they are is inductive by rank. Furthermore, the proof of the agreement of two descriptions relies on the type A theory of [14].

As for type D, Chinta and Gunnells [22] give a conjectural construction of a Weyl group multiple Dirichlet series of type D. The p-part of a series is produced as a sum over a highest weight crystal associated to an irreducible representation of $SO(2r)$. The contribution of a crystal element to the sum is described in terms of the corresponding Littelmann pattern.

We also mention constructions of Whittaker functions. McNamara [39] considers Whittaker functions on metaplectic covers of a simple algebraic group over a nonarchimedean local field. The Whittaker function is given as an integral (over the unipotent radical). Given a reduced decomposition of the long element in the Weyl group, one may break up the domain of integration into a set of cells. These cells are in a natural bijection with elements of an (infinite) crystal. By computing the integral on each cell, the Whittaker function is produced as a sum over a(n infinite) crystal structure. In type A, the resulting formula for the Whittaker function agrees with the formula given for local parts of a Weyl group multiple Dirichlet series by [15].

The averaging approach can also be used to construct metaplectic Whittaker functions as shown by Chinta and Offen [24] in the type A case, and McNamara [40] in general. Work of McNamara thereby provides a number-theoretic proof that the two methods (averaging and crystal constructions) produce the same local parts.

From a combinatorial perspective, the formulas produced by the two separate approaches (averaging or sum over a crystal) are related in the nonmetaplectic case by a theorem of Tokuyama [45]. More generally, Demazure operators can be used to elucidate the connection between the two approaches [23, 44] combinatorially.

This relies heavily on the branching properties of crystals. This is explained in more detail in Sects. 1.3 and 1.4.

The reason for the emergence of crystal bases in the study of these topics in itself warrants further exploration. Some results of this flavor exist both in the local and in the global setting. As mentioned above, the work of McNamara [39] expresses a Whittaker function as a sum over cells of the unipotent radical. The cell decomposition is then related to geometric realizations of the crystal in terms of Lusztig data [37] and Mirković–Vilonen cycles [7]. In the global setting, Brubaker and Friedberg [8] study Whittaker coefficients of metaplectic Eisenstein series induced from maximal parabolics. They produce a formula for the Whittaker coefficient for a wide class of long words in the Weyl group by matching contributions with Lusztig data through MV polytopes considered by Kamnitzer [29].

In addition, highest weight crystals are not the only combinatorial device that is of use in constructing these number-theoretic objects. An other approach uses metaplectic ice [9, 10, 13].

1.3 Tokuyama's Theorem

In this section, we explain how the results above relate to a deformation of the Weyl character formula by Tokuyama [45], and how understanding the branching structure of crystals elucidates the relationship of the constructions.

The constructions produce a polynomial $P_\lambda(\mathbf{x})$ that satisfies certain functional equations under a Weyl group W. (Here we assume that the polynomial ring $\mathbb{C}[\mathbf{x}]$ in r variables is identified with the group algebra of the weight lattice.) This is done either by taking a sum over a highest weight crystal $\mathcal{C}_\lambda$, or by taking an average over the Weyl group. We explain how both of these strategies results in a polynomial that is, roughly speaking, a deformation of a Weyl character.

First let us consider the method of producing the polynomial P_λ by taking a sum over the crystal graph:

$$P_\lambda(\mathbf{x}) = \sum_{b \in \mathcal{C}_\lambda} G(b)\mathbf{x}^{\mathrm{wt}(b)} \tag{1}$$

The elements of the crystal $\mathcal{C}_\lambda$ are in bijection (via the weight map wt) with a weight basis of a representation of highest weight λ. Note that if we had $G(b) = 1$ for every element of the crystal in (1), then the resulting sum would be the character χ_λ of this highest weight representation:

$$\sum_{b \in \mathcal{C}_\lambda} 1 \cdot \mathbf{x}^{\mathrm{wt}(b)} = \chi_\lambda(\mathbf{x}) = \frac{1}{\Delta} \cdot \sum_{w \in W} (-1)^{\ell(w)} \cdot w(\mathbf{x}^{\lambda+\rho}) \tag{2}$$

Here the right-hand side is the Weyl character formula, and Δ is a Weyl denominator.

In general $G(b)$ is more complicated, but in the simplest case we have that in fact

$$P_\lambda(\mathbf{x}) = \sum_{b \in \mathcal{C}_\lambda} G(b)\mathbf{x}^{\mathrm{wt}(b)} = \Delta_q \cdot \chi_\lambda(\mathbf{x}) \tag{3}$$

where Δ_q is a deformation of the Weyl denominator.

Next let us consider the "averaging approach" to constructing the polynomial $P_\lambda(\mathbf{x})$. This approach produces $P_\lambda(\mathbf{x})$ by an expression similar to the right-hand side of (2). However, the action of w on the monomial $\mathbf{x}^{\lambda+\rho}$ is replaced by the Chinta–Gunnells action [21], which depends on the metaplectic degree n. In the special case of $n = 1$, this construction results in the expression $\Delta_q \cdot \chi_\lambda(\mathbf{x})$ as well.

The statement that the two approaches to constructing P_λ give the same result can thus be phrased as a combinatorial identity, a deformation of the Weyl character formula. In the nonmetaplectic case, this is the second equality in (3), and this identity is a theorem by Tokuyama [45].

When $n > 1$, then understanding the relationship between the two constructions of P_λ combinatorially amounts to proving a metaplectic analogue of Tokuyama's theorem. In the type A case, this was done by the author in [44] using metaplectic Demazure–Lusztig operators defined in [23].

We mention that analogues of Tokuyama's theorem for root systems of other types have been given by Hamel and King [28] (for type B) and Friedlander, Gaudet, and Gunnells [27] (in type G_2). Note also that the agreement between the relevant constructions in the type A case follows from work of McNamara as indicated above. However, treating these sides combinatorially via Demazure–Lusztig operators allows one to understand how the constructions can be extended to more general settings, for example, from the finite dimensional to the general Kac–Moody setting [42], or, from Whittaker functions to the constructions of Iwahori–Whittaker functions [43].

The proof of the metaplectic analogue of Tokuyama's theorem in [44] relies heavily on the type A crystal construction "respecting" the branching structure of the highest weight crystal. We explain this in more detail next.

1.4 Motivation: Demazure–Lusztig Operators and the Branching Structure

As seen above, understanding the combinatorial relationship between different constructions of the polynomial P_λ (which may be a Whittaker function or the prime part of a WMDS) amounts to proving a metaplectic analogue of Tokuyama's theorem.

Using Demazure–Lusztig operators, one may phrase a more general identity, corresponding to elements w of the Weyl group, and any metaplectic degree n. The more general identity [44, Theorem 1.] is of the form:

$$\left(\sum_{u \le w} T_u\right)\mathbf{x}^\lambda = \sum_{b \in \mathcal{C}_\lambda^{(w)}} G(b)\mathbf{x}^{\mathrm{wt}(b)} \tag{4}$$

Here the expression on the left-hand side can be thought of as the general form of the expression produced by the averaging method (by results in [23]) and the right-hand side is a sum over a Demazure crystal. The "metaplectic analogue of Tokuyama's theorem" (in type A) is the special case of this statement corresponding to w being the long element of the Weyl group. This more general statement has the advantage that it can be proven "one simple reflection at a time," i.e., by induction on the length $\ell(w)$ of the Weyl group element.

The fact that the construction of P_λ as a sum over a highest weight crystal respects the branching structure of the crystal is crucial to the proof. We explain what we mean by this below. We shall return to this discussion in more detail in Sect. 5 equipped with the necessary background.

The crystal $\mathcal{C}_\lambda$ is graph, whose edges are labeled by simple roots α_i $(1 \leq i \leq r)$ of an underlying Lie algebra or rank r. When the edges labeled by α_r are omitted, the remaining graph is a disjoint union of crystals $\mathcal{C}_\mu$, corresponding to a Lie algebra of the same Cartan type as $\mathfrak{g}$, but rank $r-1$:

$$\mathcal{C}_\lambda = \sqcup_\mu \mathcal{C}_\mu \tag{5}$$

A crystal element $b \in \mathcal{C}_\lambda$ has a contribution $G(b) = G_\lambda(b)$ in the sum (1). The element b belongs to exactly one of the rank $r-1$ crystals $\mathcal{C}_\mu \subset \mathcal{C}_\lambda$. The element b has a contribution in the analogous construction of $P_\mu(\mathbf{x})$. By the constructions respecting the branching structure, we mean that we have

$$G_\lambda(b) = g(\mu) \cdot G_\mu(b), \tag{6}$$

where the factor $g(\mu)$ is the same for every element $b \in \mathcal{C}_\mu \subset \mathcal{C}_\lambda$. It follows that P_λ can be written as an expression of polynomials P_μ corresponding to the weights μ of the decomposition (5) above.

This means that statements about these crystal constructions are amenable to proof by induction on rank. The parametrization of crystal elements by Littelmann patterns highlights the branching structure of crystals. We encourage the reader to keep the branching structure in mind while reading through the sections covering the ingredients of the constructions.

2 Preliminaries

Before describing the constructions mentioned above, we cover a few preliminaries. The constructions in Sect. 4 have two main ingredients: a set of root data and an arithmetic ingredient in the form of certain Gauss sums. We introduce notation and the necessary ingredients below.

2.1 Notation

Throughout the paper, Φ shall denote a root system of rank r, with Φ^+ (respectively Φ^-) being the set of positive (respectively, negative) roots. Let $\Delta = \{\alpha_1, \ldots, \alpha_r\}$ be a set of simple roots in Φ, and let us write $\rho = \frac{1}{2}\sum_{\beta\in\Phi^+}\beta$ for the Weyl vector. Of particular interest are the root systems of Cartan types A, B, C and D. We give [6] as a general reference on root systems. Note however that when discussing Littelmann patterns [34], our numbering of the simple roots agrees with that of Littelmann and hence differs from that of Bourbaki. (For example in type D, the simple roots α_1 and α_2 are orthogonal.)

Let us write $\sigma_1, \ldots, \sigma_r$ for the set of simple reflections corresponding to the simple roots; σ_i is the reflection through the hyperplane perpendicular to α_i. The Weyl group W is generated by the simple reflections σ_i $(1 \leq i \leq r)$. Every element $w \in W$ can be written as a product $w = \sigma_{i_1}\cdots\sigma_{i_k}$. We call this a reduced decomposition and $\underline{w} = [i_1, \ldots, i_k]$ a reduced word if k is minimal and $k = \ell(w)$, the length of w. The Weyl group has a unique longest element $w_0 \in W$. The parameterization of highest weight crystals by Littelmann patterns given in Sect. 3 depends on a choice of a *nice decomposition* $\underline{w}_0$ of the long element.

The Weyl group permutes the elements of Φ. Let $\Phi(w) = w^{-1}(\Phi^-)\cap\Phi^+$, then $\ell(w) = |\Phi(w)|$ and $\Phi(w_0) = \Phi^+$. We shall denote the weight lattice corresponding to Φ by Λ, and the fundamental weights corresponding to the basis Δ by $\varpi_1, \ldots, \varpi_r$. The constructions we are concerned with produce polynomials in $\mathbb{C}[\Lambda]$. The Weyl group has a natural action on Λ, and hence on $\mathbb{C}[\Lambda]$, it thus makes sense to talk about the resulting polynomials being symmetric (or having functional equations) under the Weyl group.

2.2 Gauss Sums

Next we introduce notation for the arithmetic ingredients of the polynomials constructed in Sect. 4. The contribution of a crystal element is given by nth-order Gauss sums, where n is a positive integer. (In applications, n is the degree of the metaplectic cover.) Recall that (for $n = 2$), one may take the quadratic Gauss sum

$$G(a) = \sum_{k=0}^{p}\left(\frac{k}{p}\right)e^{\frac{ka2\pi i}{p}} \tag{7}$$

where $\left(\frac{k}{p}\right)$ is the Legendre symbol; i.e., it is 1 if k is a square modulo p and -1 otherwise. The Gauss sums appearing in the construction are generalizations of the one in (7). The Legendre symbol is replaced by an nth power residue symbol (a multiplicative character), and $e^{\frac{ka2\pi i}{p}}$ is replaced by an additive character.

To state the definition of the Gauss sums $g(a)(=g_t(a))$ and $h(a)(=h_t(a))$, we introduce some more notation. We use the language of local and global fields, and provide examples (see [41] for reference). The Gauss sums $g(a)$ and $h(a)$ are functions that depend on the residue of a modulo n. For the purpose of understanding the constructions and their relationship to the branching structure of crystals, the values of these functions are not crucial.

Following [11], let F be a global field; the reader may choose to simply think of $\mathbb{Q}$ as an example. For a place v of F, one may take the completion F_v. (For example, the completion $\mathbb{Q}_p$ of p-adic numbers at any finite prime p, or the completion $\mathbb{R}$ at the infinite place.) Let $\mathcal{O}_v$ denote the set of integers (e.g., $\mathbb{Z}_p \subset \mathbb{Q}_p$ or $\mathbb{Z} \subset \mathbb{R}$). Let S be a finite set of places of F, and let $\mathcal{O}_S$ denote the set of S-integers $x \in K$ such that $x \in \mathcal{O}_v$ for every $v \notin S$. (For $S = \{\infty, 2\}$, the set $\mathcal{O}_S \subseteq \mathbb{Q}$ is the set of rational numbers with only 2 in the denominator.) For a sufficiently large S, $\mathcal{O}_S$ is a principal ideal domain. Let $F_S = \prod_{v\in S} F_v$, $\mathcal{O}_S$ embeds into F_S diagonally. Let ψ be a character of F_S trivial on $\mathcal{O}_S$ and no larger fractional ideal. Let $\left(\frac{\cdot}{\cdot}\right)_n$ denote the nth-order residue symbol and t a positive integer. We define

$$g_t(a,c) = \sum_{d \mod c} \left(\frac{d}{c}\right)_n^t \psi\left(\frac{ad}{c}\right). \tag{8}$$

The constructions in Sect. 4 will involve special values of $g_t(a,c)$. We shall have $t = 1$ or $t = 2$ be the length of a simple root (i.e., $t = 1$ in the simply laced cases and $t = 1$ or $t = 2$ in type B or type C), and we shall have fixed a prime p. Then we set

$$g_t(a) = g_t(p^{a-1}, p^a); \text{ and } h_t(a) = g_t(p^a, p^a) = \begin{cases} |(\mathcal{O}_S/p\mathcal{O}_S)^\times| & \text{if } t^{-1}n \mid a \\ 0 & \text{if } t^{-1}n \nmid a \end{cases} \tag{9}$$

In the remainder of the paper, we use the notation $q = |\mathcal{O}_S/p\mathcal{O}_S|$ for the order of a residue field.

The polynomials P_λ that we shall define in Sect. 4 are given as a sum over a crystal. Each term is determined via combinatorial data coming from the parameterization of the corresponding crystal element via a method that makes use of the above Gauss sums. For example in Cartan type A, the Gauss sums take n values (indexed by the residue classes modulo n). Consequently, any statement about these polynomials can be phrased entirely in terms of the structure of the highest weight crystal and identities of these nth-order Gauss sums. Such identities are rare. In addition to the identity expressing the relationship of Gauss sums corresponding to conjugate characters, one has the Hasse–Davenport relations [25]. By work of Yamamoto [46], these are essentially the only multiplicative identities of these Gauss sums. We thank one of the referees for pointing this out.

2.3 Highest Weight Crystals and Littelmann's Cone

Given an irreducible (finite) root system Φ and a dominant weight λ, there is an associated crystal graph $\mathcal{C}_\lambda$. We shall describe the structure of $\mathcal{C}_\lambda$ as a directed graph with colored edges. We mention that if $\mathfrak{g}$ is a simple Lie algebra with root system Φ, and V_λ is the unique simple $\mathfrak{g}$ module with highest weight λ, then the quantized universal enveloping algebra $U_{\mathsf{q}}(\mathfrak{g})$ has a corresponding module. A crystal base is a base for this module at $\mathsf{q} = 0$. It carries a graph structure induced by its structure as a $U_{\mathsf{q}}(\mathfrak{g})$ module. For further information, see [30]. Here we forgo exploring the connection with the quantum group. We instead explain the structure of a crystal as a colored directed graph and the parameterization of crystals by Berenstein–Zelevinsky–Littelmann paths and Littelmann patterns.

2.3.1 The Crystal as a Colored Directed Graph

We now describe $\mathcal{C}_\lambda$ as a colored directed graph. Let B be a finite set, and we call elements of B elements (or vertices) of the crystal. (We shall abuse notation and write $b \in \mathcal{C}_\lambda$ for a $b \in B$.) For every $1 \leq i \leq r$, we have operators $f_i : B \cup \{0\} \to B \cup \{0\}$ and $e_i : B \sqcup \{0\} \to B \sqcup \{0\}$ acting on the vertices. We shall refer to these as root operators. They have the property that if $b, b' \in B$, then $f_i b = b'$ and $b = e_i b'$ are equivalent. This defines the structure of $\mathcal{C}_\lambda$ as a colored directed graph: if $b, b' \in B$ and $f_i b = b'$, then $\mathcal{C}_\lambda$ has a directed edge $b \xrightarrow{i} b'$ "colored" by the index i. There is a weight function $\mathrm{wt} : B \to \Lambda$ such that $\mathrm{wt}(f_i(b)) = \mathrm{wt}(b) - \alpha_i$ and in fact the function wt is a bijection between B and a weight basis of the highest weight $\mathfrak{g}$-module V_λ. In particular, there is a unique "highest element" $b_\lambda \in \mathcal{C}_\lambda$ with $\mathrm{wt}(b_\lambda) = \lambda$. This b_λ is the unique element of B such that $e_i b_\lambda = 0$ for every $1 \leq i \leq r$. It follows that

$$B \sqcup \{0\} = \{f_{i_1}^{n_1} f_{i_2}^{n_2} \cdots f_{i_k}^{n_k} b_\lambda \mid 1 \leq i_j \leq k, 0 \leq n_j\}. \tag{10}$$

We shall be interested in writing an element $b \in B$ as $b = f_{i_1}^{n_1} f_{i_2}^{n_2} \cdots f_{i_k}^{n_k} b_\lambda$ in particular when the sequence of indices $[i_1, i_2 \ldots, i_k]$ is a reduced word.

2.3.2 Berenstein–Zelevinsky–Littelmann Paths

Let $\underline{w} = [i_1, i_2 \ldots, i_k]$ be a reduced word in W, and let $\mathfrak{n} = [n_1, n_2 \ldots, n_k] \in (\mathbb{Z}_{\geq 0})^k$ for $b = f_{i_1}^{n_1} f_{i_2}^{n_1} \cdots f_{i_k}^{n_k} b_\lambda$. We call $\mathfrak{n}$ an adaptive string of b [34] if for every $1 \leq j \leq k$ we have

$$1 \leq j \leq k : \; e_{i_j} f_{i_{j+1}}^{n_1} \cdots f_{i_k}^{n_k} b_\lambda = 0 \tag{11}$$

We can think of $\mathfrak{n}$ as encoding a path from b to b_λ along crystal edges (against the direction of the edges), using $\underline{w}$ as a road map. To get the path, starting at b we first take steps along edges colored i_1 as long as that is possible. After taking n_1 steps, we

arrive at a vertex $b_1 = e_{i_1}^{n_1} b$ such that $e_{i_1} b_1 = 0$. We then proceed with steps along edges colored i_2 for as long as possible, etc.

Taking an adaptive string above defines a map $b \mapsto \mathbb{Z}_{\geq 0}{}^{\ell(\underline{w})}$ for any reduced word $\underline{w}$. Let $\underline{w}_0$ be a long word of the Weyl group. Write $S_{\underline{w}_0}^{\lambda} \subseteq \mathbb{Z}_{\geq 0}{}^{\ell(\underline{w}_0)}$ for the set of adaptive strings that occur in $\mathcal{C}_\lambda$ and $S_{\underline{w}_0} \subseteq \mathbb{Z}_{\geq 0}{}^{\ell(\underline{w}_0)}$ for the set of strings that occur for any strongly dominant λ. Then it follows from work of Littelmann, Berenstein, and Zelevinsky [3, 33, 34] that $S_{\underline{w}_0}$ is the set of integral points inside a convex cone, which we from now on refer to as the *Littelmann cone* $C_{\underline{w}_0}$. Furthermore, the set $S_{\underline{w}_0}^{\lambda}$ is the set of integral points in a convex polytope $C_{\underline{w}_0}^{\lambda}$ in this cone (the *Littelmann polytope*). The inequalities describing $C_{\underline{w}_0}$ depend on the long word $\underline{w}_0$; the further inequalities describing $C_{\underline{w}_0}^{\lambda}$ depend on λ as well. For particularly "nice" choices of $\underline{w}_0$ [34], these inequalities take on a transparent form. We shall describe these choices for Cartan types A, B, C, and D as well as the Littelmann patterns they give rise to in Sect. 3. For $\underline{w} = \underline{w}_0$, we shall refer to the adaptive string $\mathfrak{n}$ corresponding to a vertex $b \in \mathcal{C}_\lambda$ (as well as the corresponding path in $\mathcal{C}_\lambda$) as the Berenstein–Zelevinsky–Littelmann path or BZL path of b and write $BZL(b) = \mathfrak{n}$.

2.4 Multiple Dirichlet Series and Whittaker Functions

We briefly introduce the objects from number theory that are produced by the constructions in Sect. 4. Since we wish to focus on the combinatorics of the constructions, we keep the length of this section to a minimum. Our purpose here is merely to motivate the appearance of highest weight crystals as an apt combinatorial device in the study of these objects.

2.4.1 Multiple Dirichlet Series

Multiple Dirichlet series are series in several complex variables. They can be used to study automorphic L-functions, generalizations of the Riemann zeta function via the Langlands–Shahidi method. Of special interest to us here are Weyl group multiple Dirichlet series, whose functional equations are governed by a Weyl group associated to a (finite) Cartan type. The functional equations are of significance in proving meromorphic continuation and functional equations. We explain briefly how prime power coefficients of multiple Dirichlet series are related to sums over a highest weight crystal. We follow the notation of [17] with some simplifications, so as not to occlude the picture.

Let F now be a global field. We wish to construct a series in r variables $s_1, \dots, s_r$:

$$\sum_{C_i} H(C_1, \dots, C_r; m_1, \dots, m_r) \cdot |C_1|^{-2s_1} \cdots |C_r|^{-2s_r} \tag{12}$$

where the summation is over ideals C_i of $\mathcal{O}_S$. Relating such a series to automorphic L-functions imposes certain restrictions on its construction. For example, though a series does not have an Euler product in the way the Riemann zeta function does:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}}$$

its coefficients satisfy a twisted multiplicativity and the series is hence determined by its p-parts

$$\sum_{k_i=1}^{\infty} H(p^{k_1}, \dots, p^{k_r}; p^{l_1}, \dots, p^{l_r}) \cdot |p|^{-2k_1 s_1 - \dots - 2k_r s_r} \tag{13}$$

where p is (a representative of) a prime ideal, and $(l_1, \dots, l_r)$ correspond to a weight $\lambda = \sum_{i=1}^{r} l_i \varpi_i$.

Constructing a Weyl group multiple Dirichlet series thus amounts to describing the coefficients $H(p^{k_1}, \dots, p^{k_r}; p^{l_1}, \dots, p^{l_r})$ for any fixed weight λ. Note that assigning a weight to every $(k_1, \dots, k_r)$ as above, we may interpret the p-part as a sum over the weight lattice Λ. Its support turns out to be finite, and in fact contained in the convex hull of the Weyl group orbit of λ.

Recall that for a highest weight crystal $\mathcal{C}_\lambda$ (associated to a root system Φ), the weight function $\mathrm{wt} : \mathcal{C}_\lambda \to \Lambda$ is a bijection between vertices of $\mathcal{C}_\lambda$ and a weight basis of a representation with highest weight λ. Hence the constructions of the p-part may be written as a sum over a highest weight crystal.

2.4.2 Whittaker Functions

Our aim here is to motivate why metaplectic analogues of the Casselman–Shalika formula lead to constructions involving highest weight crystals.

Let $\mathbf{G}$ be a split reductive group defined over $\mathbb{Z}$. (The reader may think of SL_r or Sp_{2r}.) Let F be a nonarchimedean local field (e.g., $F = \mathbb{Q}_p$, the p-adic numbers), and $\mathcal{O} \subset F$ the ring of integers in F (e.g., $\mathcal{O} = \mathbb{Z}_p$). Let $G = \mathbf{G}(F)$ and $K = \mathbf{G}(\mathcal{O})$ be a maximal compact in G. Let $T \subseteq G$ be a maximal torus, and $U \subset G$ be the unipotent radical of a Borel subgroup of G. (In the examples above, T is the group of diagonal matrices in G and U the group of upper triangular matrices with 1s on the diagonal.) Let $\widehat{G}$ denote the Langlands dual of G (we have $\widehat{\mathrm{SL}}_{r+1} = \mathrm{PGL}_r$ and $\widehat{\mathrm{Sp}}_{2r} = \mathrm{SO}_{2r+1}$); let Φ be the root system associated with $\widehat{G}$ and Λ its weight lattice. (Here Φ is of type A or type B for SL_r or Sp_{2r}, respectively.) To an element $\mathbf{x} \in \widehat{T}$, we may associate a Whittaker function $\mathcal{W} : G \to \mathbb{C}$ that satisfies $\mathcal{W}(ugk) = \psi(u)\mathcal{W}(g)$ (for $u \in U$, $g \in G$, and $k \in K$, where ψ is an unramified character of U). Let $\pi \in F$ be a uniformizer (e.g., p in $F = \mathbb{Q}_p$). By the Iwasawa decomposition, any element $g \in G$ can be written as $g = u\pi^\lambda k$, where $u \in U$, $k \in K$, and $\lambda \in \Lambda$ is a cocharacter

of T. A Whittaker function $\mathcal{W}$ is then determined by its values $\mathcal{W}(\pi^\lambda)$. In this classical (nonmetaplectic) setting, these values are determined by the Casselman–Shalika formula [19]:

$$\mathcal{W}(\mathbf{x}, \lambda) = \prod_{\alpha\in\Phi^+}(1 - q^{-1}\mathbf{x}^\alpha)\chi_\lambda(\mathbf{x}) \tag{14}$$

This expresses the values of a Whittaker function $\mathcal{W}(\mathbf{x}, \lambda)$ in terms of the character $\chi_\lambda(\mathbf{x})$ of a representation of $\widehat{G}$ of highest weight λ. (Here $q = |\mathcal{O}/\pi\mathcal{O}|$ as before.)

Now let n be a positive integer so that $\mathrm{char}F \nmid n$ and $|\mu_{2n}| = 2n$ for the group $\mu_{2n} \subset F$ of nth roots of unity. Then an n-fold metaplectic cover $\tilde{G}$ of G is a central extension

$$1 \to \mu_n \to \tilde{G} \to G \to 1$$

constructed from the Cartan datum of G and some arithmetic data on F [38]. By a metaplectic generalization of the Casselman–Shalika formula, we mean an analogue of (14) for Whittaker functions on $\tilde{G}$. As mentioned in Sect. 1.3, such a generalization may produce such a formula as a sum over a highest weight crystal, or as a sum over a Weyl group. This is motivated by the shape of (14) and its similarity with the deformation of the Weyl character formula in (3).

3 Littelmann Patterns

We recall Littelmann patterns from [34] in each of the Cartan types A_r, B_r, C_r, and D_r. A pattern is an array of $\ell(w_0)$ nonnegative integers. Integral points of the Littelmann cone (see Sect. 2.3) are in bijection with the set of patterns that satisfy a set of inequalities. Imposing a further set of inequalities gives a parametrization of integral points within the Littelmann polytope, i.e., a highest weight crystal for a fixed highest weight. The contribution of a single element to the sums in Sect. 4 will be phrased in terms of the corresponding Littelmann pattern.

The branching properties of highest weight crystals and how it is reflected in the constructions will be made explicit in Sect. 5. One may observe these branching properties in the extent to which the Littelmann patterns are consistent within an infinite family of Cartan types. Note also that the simple root α_r that is "new" in rank r is associated only to entries in the top row of the pattern.

3.1 The Choice of a Long Words

Recall that a long word is a reduced decomposition of the long element of the Weyl group. The parametrization of crystal elements in terms of Littelmann patterns is dependent on the choice of a long word $\underline{w}_0$. The choice of particular "nice" long

words results in the Littelmann cone having a transparent description. We give the nice long words here for each infinite family of Cartan types.

Notice that the choice is consistent within each family in the following sense. Let X stand for any of A, B, C, or D, and let $\underline{w}_0^{X_r}$ be the choice of long word for type X_r, i.e., rank r. Then the word $\underline{w}_0^{X_r}$ starts with the long word $\underline{w}_0^{X_{r-1}}$ from rank $r-1$.

The choices are as follows.

$$\underline{w}_0^{A_r} = [(1), (2, 1), (3, 2, 1), \ldots, (r, r-1, \ldots, 2, 1)] \tag{15}$$
$$\underline{w}_0^{B_r} = \underline{w}_0^{C_r} = [(1), (2, 1, 2), \ldots, (r, r-1, \ldots, 2, 1, 2, \ldots, r)] \tag{16}$$
$$\underline{w}_0^{D_r} = [(1), (2), (3, 1, 2, 3), \ldots, (r, r-1, \ldots, 3, 1, 2, 3, \ldots, r)] \tag{17}$$

3.2 *The Shape of Patterns*

The choice of a long word $\underline{w}_0$ establishes a bijection between elements of a crystal and $\ell(\underline{w}_0)$-tuples of nonnegative integers via BZL paths as in Sect. 2.3.2. We arrange these $\ell(w_0)$ integers as entries $a_{i,j}$ of a Littelmann pattern. The shape of the pattern reflects the choice of $\underline{w}_0$ made.

Each column of a pattern corresponds to a particular index $1 \leq j \leq r$. Entries $a_{i,j}$ with the same column index j correspond to occurrences of same simple reflection in the word $\underline{w}_0$. A row of the pattern will correspond to a step in the rank within the infinite family of Cartan types.

In the remainder of this chapter, we follow the convention that if $a_{i,j}$ is not an entry of a pattern, then $a_{i,j} = 0$. (This is the case, for example, if $i \leq 0$ or $j < i$.)

3.2.1 Type A_r

We have $\ell(\underline{w}_0^{A_r}) - \ell(\underline{w}_0^{A_{r-1}}) = r$ for $r \geq 2$. A Littelmann pattern of type A_r has r rows, with $r - i + 1$ elements in the ith row. We write $\mathfrak{L} = (a_{i,j})_{\substack{1\leq i\leq r\\ i\leq j\leq r}}$ and draw the pattern aligned to the right:

$$\begin{array}{|c|c|c|c|} \hline a_{1,1} & a_{1,2} & \cdots & a_{1,r} \\ \hline \multicolumn{1}{c|}{} & a_{2,2} & \cdots & a_{2,r} \\ \cline{2-4} \multicolumn{1}{c}{} & \multicolumn{1}{c|}{} & \ddots & \vdots \\ \cline{3-4} \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c|}{} & a_{r,r} \\ \cline{4-4} \end{array} \tag{18}$$

3.2.2 Type B_r and C_r

In this case, $\ell\left(\underline{w}_0^{B_r}\right) - \ell\left(\underline{w}_0^{B_{r-1}}\right) = 2r - 1$. These Littelmann patterns have r rows as well, but now the ith row has $2r - 1$ entries, denoted $a_{i,j}$ for $i \leq j \leq 2r - i$. We write $\bar{j} = 2r - j$ and $\bar{a}_{i,j} = a_{i,\bar{j}}$, and draw the patterns centered as follows:

$$
\begin{array}{|c|c|c|c|c|c|c|}
\hline
a_{1,1} & a_{1,2} & \cdots & a_{1,r} & \cdots & \bar{a}_{1,2} & \bar{a}_{1,1} \\
\hline
\multicolumn{1}{c|}{} & a_{2,2} & \cdots & a_{2,r} & \cdots & \bar{a}_{2,2} & \multicolumn{1}{c}{} \\
\cline{2-6}
\multicolumn{1}{c}{} & \multicolumn{1}{c|}{} & \ddots & \vdots & \iddots & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} \\
\cline{3-5}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c|}{} & a_{r,r} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} \\
\cline{4-4}
\end{array}
\tag{19}
$$

3.2.3 Type D_r

In this case, $\ell\left(\underline{w}_0^{D_r}\right) - \ell\left(\underline{w}_0^{D_{r-1}}\right) = 2r - 2$ for $r \geq 3$. The $\ell\left(\underline{w}_0^{D_r}\right) = r^2 - r$ integers from a BZL path are now arranged into a Littelmann pattern with $r - 1$ rows. The ith row has $2r - 2i$ entries, $a_{i,j}$ for $i \leq j \leq 2r - 1 - i$. We use notation similar to type B and C and write $\bar{j} = 2r - 1 - j$ for $\bar{a}_{i,j} = a_{i,2r-1-j}$.

$$
\begin{array}{|c|c|c|c|c|c|c|c|}
\hline
a_{1,1} & a_{1,2} & \cdots & a_{1,\bar{r}} & a_{1,r} & \cdots & \bar{a}_{1,2} & \bar{a}_{1,1} \\
\hline
\multicolumn{1}{c|}{} & a_{2,2} & \cdots & a_{2,\bar{r}} & a_{2,r} & \cdots & \bar{a}_{2,2} & \multicolumn{1}{c}{} \\
\cline{2-7}
\multicolumn{1}{c}{} & \multicolumn{1}{c|}{} & \ddots & \vdots & \vdots & \iddots & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} \\
\cline{3-6}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c|}{} & a_{\bar{r},\bar{r}} & a_{\bar{r},r} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} \\
\cline{4-5}
\end{array}
\tag{20}
$$

3.3 The Bijection with Crystal Elements

We are ready to give the bijection between crystal elements and Littelmann patterns.

Recall that the BZL path of a crystal element b consists of $\ell(\underline{w}_0)$ segments. Taking the length of these segments produces a tuple $BZL(b) = (n_1, \ldots, n_{\ell(\underline{w}_0)})$. The entries of the Littelmann pattern $\mathfrak{L}(b)$ corresponding to b are these integers n_h ($1 \leq h \leq \ell(\underline{w}_0)$). The pattern $\mathfrak{L}(b)$ is filled with elements of $BZL(b)$ row by row proceeding from left to right and from bottom to top. For example, in type A_r we have that $\mathfrak{L}(b) = (a_{i,j})_{\substack{1 \leq i \leq r \\ i \leq j \leq r}}$ and:

$$
a_{r,r} = n_1,\ a_{r-1,r-1} = n_2,\ a_{r-1,r} = n_3, \ldots,\ a_{1,1} = n_{\binom{r}{2}+1}, \ldots,\ a_{1,r} = n_{\binom{r+1}{2}}
$$

The shape of the Littelmann patterns above arranges entries in the same column if they correspond to the same edge label. We examine this property in more detail.

3.4 *The Weight of a Pattern*

Let b be an element in a crystal element of highest weight λ. Let $BZL(b) = (n_1, \ldots, n_{\ell(\underline{w}_0)})$ and $\mathfrak{L} = \mathfrak{L}(b)$ be the Littelmann pattern corresponding to b via the bijection above. Then the weight of b is easily recovered from entries of the pattern. Recall that the hth segment of the BZL path follows edges of the crystal labeled with index $k = \underline{w}_0(h)$. These edges all correspond to a root operator for the simple root α_k; i.e., they all have the same label k. It follows that

$$\lambda - \mathrm{wt}(b) = \sum_{k=1}^{r} \alpha_k \cdot \sum_{\underline{w}_0(h)=k} n_h. \tag{21}$$

The shape of the patterns has the following property. Entries in a single column of $\mathfrak{L}(b)$ correspond to segments of the BZL path of b. These segments all run along edges of the crystal with the same color k (or α_k) $1 \leq k \leq r$. Figure 1 shows the index of the crystal edges corresponding to each column in the various types. Observe that reading off the index for elements in the top row gives the segment of $\underline{w}_0$ that is present in rank r but not in rank $r-1$.

We make this explicit for each of the infinite families. We define the weight $\mathfrak{s}(\mathfrak{L}) = (s_1, \ldots, s_r)$ of a Littelmann pattern $\mathfrak{L}$ ($s_k = s_k(\mathfrak{L})$) so that:

$$\lambda - \mathrm{wt}(b) = \sum_{k=1}^{r} s_k(\mathfrak{L}(b)) \cdot \alpha_k \tag{22}$$

3.4.1 Type A

In this case, $\mathfrak{L}(b) = (a_{i,j})_{1 \leq i \leq j \leq r}$. A column consists of entries $a_{1,j}, \ldots, a_{j,j}$. If $BZL(b) = (n_1, \ldots, n_{\binom{r+1}{2}})$, then $a_{i,j} = n_{\binom{r-i}{2}+j-i}$ and the corresponding segment of the BZL path of b lies along edges labeled $r - j + 1$. We define $s_k(\mathfrak{L})$ for $1 \leq k \leq r$ by:

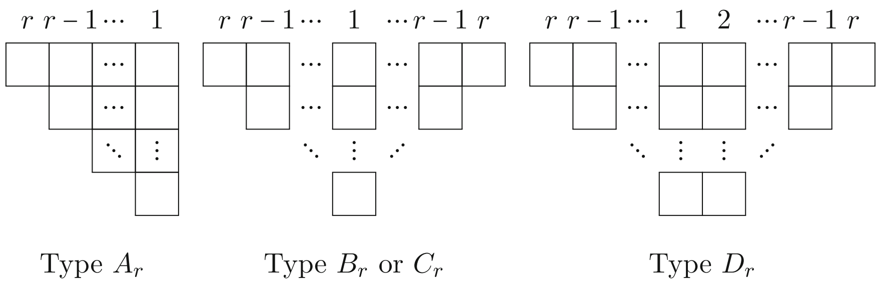


Fig. 1 Edge colors corresponding to columns of a pattern

$$s_k(\mathfrak{L}) = \sum_{i=1}^{r} a_{i,r+1-k} \tag{23}$$

3.4.2 Type B or C

Here we have $\mathfrak{L}(b) = (a_{i,j})_{1\le i\le j\le 2r-i}$. For any $1 < k \le r$ there are two columns corresponding to the edge index k, the one with $j = r - k + 1$ and the one with $\bar{j} = r - k + 1$. We thus define:

$$s_k(\mathfrak{L}) = \sum_{i=1}^{r}(a_{i,r+1-k} + \bar{a}_{i,r+1-k}) \tag{24}$$

The middle column corresponds to the index 1 and so we define:

$$s_1(\mathfrak{L}) = \sum_{i=1}^{r} a_{i,r} \tag{25}$$

Note that $|\alpha_1|$ is different from $|\alpha_2| = \cdots = |\alpha_r|$.

3.4.3 Type D

This case is similar to the previous one. We have $\mathfrak{L}(b) = (a_{i,j})_{1\le i\le j\le 2r-1-i}$. For $2 < k \le r$, the two columns corresponding to the edge index k are the jth where $j = r - k + 1$ and $\bar{j}$th, where recall that $\bar{j} = r - k$. The two middle columns correspond to α_1 and α_2, the roots on the "branched" end of the Dynkin diagram. Hence we define (cf. [22]):

$$s_k(\mathfrak{L}) = \begin{cases} \sum_{i=1}^{r}(a_{i,r+1-k} + \bar{a}_{i,r+1-k}) & \text{if } 2 < k \le r \\ \sum_{i=1}^{r} a_{i,r-2+k} & \text{if } k = 1, 2 \end{cases} \tag{26}$$

3.5 Constraints on Littelmann Patterns

The correspondence $b \mapsto \mathfrak{L}(b)$ described in Sect. 3.3 above is a bijection between integral points of the Littelmann cone (see Sect. 2.3.2) and the set of Littelmann patterns whose entries satisfy a certain set on inequalities, depending on the Cartan type of the underlying root system. To get a set of patterns in bijection with the integral points of a Littelmann polytope for λ (equivalently, a crystal of highest weight λ), we may impose a further set of inequalities on the entries. This second set of inequalities shall depend on the highest weight λ. In this section, we make these

constraints explicit for each of the infinite families of Cartan types. The constructions in Sect. 4 phrase the contribution of a crystal element b in terms of whether these inequalities are satisfied by the entries of $\mathfrak{L}(b)$ strictly or with an equality.

3.5.1 Constraints for the Cone

We give the inequalities describing Littelmann patterns corresponding to integral points of the Littelmann cone C. Let C^{X_r} denote the Littelmann cone in the Cartan type X_r. Then we have the following.

Theorem. [34] *Let b correspond to $\mathfrak{L}(b)$ under the bijection described in Sect. 3.3. Then b is an integral point of C^{X_r} if and only if the entries of $\mathfrak{L}(b)$ are nonnegative and the following holds.*

$X_r = A_r$: [34, Theorem 5.1] *The rows are weakly decreasing:*

$$a_{i,i} \geq a_{i,i+1} \geq \cdots \geq a_{i,r-1} \geq a_{i,r} \text{ for every } 1 \leq i \leq r \tag{27}$$

$X_r = B_r$: [34, Theorem 6.1] *For every row we have:*

$$2a_{i,i} \geq 2a_{i,i+1} \geq \cdots \geq 2a_{i,r-1} \geq a_{i,r} \geq 2\bar{a}_{i,r-1} \geq \cdots \geq 2\bar{a}_{i,i} \text{ for every } 1 \leq i \leq r \tag{28}$$

$X_r = C_r$: [34, Theorem 6.1] *The rows are weakly decreasing:*

$$a_{i,i} \geq a_{i,i+1} \geq \cdots \geq a_{i,r-1} \geq a_{i,r} \geq \bar{a}_{i,r-1} \geq \cdots \geq \bar{a}_{i,i} \text{ for every } 1 \leq i \leq r \tag{29}$$

$X_r = D_r$: [34, Theorem 7.1] *For every row we have:*

$$a_{i,i} \geq a_{i,i+1} \geq \cdots \geq a_{i,r-2} \geq a_{i,r-1}, a_{i,r} \geq \bar{a}_{i,r-2} \geq \cdots \geq \bar{a}_{i,i} \text{ for every } 1 \leq i \leq r-1 \tag{30}$$

i.e., the rows are weakly decreasing with the exception of the central two elements. There is no restriction on the comparative size of these two elements.

3.5.2 Constraints for a Polytope

We introduce shorthand to refer to the sums of particular groups of elements of a Littelmann pattern. The notation $s_{i,j}(\mathfrak{L})$, $\bar{s}_{i,j}(\mathfrak{L})$, $t_{i,r-1}(\mathfrak{L})$, $t_{i,r}(\mathfrak{L})$ used here differs slightly from that of [34] ($s(a_{i,j})$, $s(\bar{a}_{i,j})$, etc.) to emphasize that $s_{i,j}(\mathfrak{L})$ may be nonzero even if $a_{i,j}$ is not an element of the pattern $\mathfrak{L}$. When the pattern $\mathfrak{L}$ is clear from context, we write $s_{i,j}$ for $s_{i,j}(\mathfrak{L})$.

We define the following shorthand:

$$s_{i,j}(\mathfrak{L}) := \begin{cases} \sum_{k=1}^{i} a_{k,j} & \text{if } \mathfrak{L} \text{ is type } A \\ \sum_{k=1}^{i} (a_{k,j} + \bar{a}_{k,j}) & \text{if } \mathfrak{L} \text{ is type } B \text{ or } C,\ j \leq r-1 \\ \sum_{k=1}^{i} a_{k,r} & \text{if } \mathfrak{L} \text{ is type } B \text{ and } j = r \\ \sum_{k=1}^{i} 2a_{k,r} & \text{if } \mathfrak{L} \text{ is type } C \text{ and } j = r \\ \sum_{k=1}^{i} (a_{k,j} + \bar{a}_{k,j}) & \text{if } \mathfrak{L} \text{ is type } D,\ j \leq r-2 \\ \sum_{k=1}^{i} (a_{k,r-1} + \bar{a}_{k,r}) & \text{if } \mathfrak{L} \text{ is type } D,\ j = r-1 \text{ or } j = r \end{cases} \tag{31}$$

$$\bar{s}_{i,j}(\mathfrak{L}) := \begin{cases} \bar{a}_{i,j} + s_{i-1,j}(\mathfrak{L}) & \text{if } \mathfrak{L} \text{ is type } B \text{ and } j \leq r-1 \\ & \text{or } \mathfrak{L} \text{ is type } C \text{ and } j \leq r \\ & \text{or } \mathfrak{L} \text{ is type } D \text{ and } j \leq r-2 \\ \bar{a}_{i,j} + 2s_{i-1,j}(\mathfrak{L}) & \text{if } \mathfrak{L} \text{ is type } B \text{ and } j = r \\ s_{i,j}(\mathfrak{L}) & \text{if } \mathfrak{L} \text{ is type } D,\ j = r-1 \text{ or } j = r \end{cases} \tag{32}$$

Observe that $\bar{s}_{i,j} = s_{i,\bar{j}}$ when both are defined; when only one is, we use this to extend the definition. For patterns $\mathfrak{L}$ of type D, we shall also need:

$$t_{i,r-1}(\mathfrak{L}) := \sum_{k=1}^{i} a_{k,r-1} \text{ and } t_{i,r}(\mathfrak{L}) := \sum_{k=1}^{i} a_{k,r} \tag{33}$$

We are ready to state the inequalities characterizing patterns that correspond to the integral points of a Littelmann polytope, or, equivalently, the points of a crystal of highest weight λ. Let $\lambda = \sum_{k=1}^{r} m_i \cdot \varpi_i$. The integers m_k appear in the inequalities.

For type A_r, the pattern $\mathfrak{L}$ corresponds to $b \in \mathcal{C}_\lambda$ if the following inequalities are satisfied [34, Corollary 4]:

$$a_{i,j} \leq m_{r-j+1} + s_{i,j-1}(\mathfrak{L}) - 2s_{i-1,j}(\mathfrak{L}) + s_{i-1,j+1}(\mathfrak{L});\ \text{ for } 1 \leq i \leq j \leq r \tag{34}$$

For type B_r and type C_r the inequalities are as follows [34, Corollary 6.]:

$$\bar{a}_{i,j} \leq m_{r-j+1} + \bar{s}_{i,j-1}(\mathfrak{L}) - 2s_{i-1,j}(\mathfrak{L}) + s_{i-1,j+1}(\mathfrak{L});\ \text{ for } 1 \leq i \leq j \leq r-1 \tag{35}$$

$$a_{i,j} \leq m_{r-j+1} + \bar{s}_{i,j-1}(\mathfrak{L}) - 2\bar{s}_{i,j}(\mathfrak{L}) + s_{i,j+1}(\mathfrak{L});\ \text{ for } 1 \leq i \leq j \leq r-1 \tag{36}$$

$$a_{i,r} \leq m_1 + d\bar{s}_{i,r-1}(\mathfrak{L}) - d\bar{s}_{i-1,r}(\mathfrak{L}) \tag{37}$$

where $d = 2$ in type B and $d = 2$ in type C.

Finally for type D_r the inequalities are as follows [34, Corollary 8.]:

$$\bar{a}_{i,j} \leq m_{r-j+1} + \bar{s}_{i,j-1}(\mathfrak{L}) - 2s_{i-1,j}(\mathfrak{L}) + s_{i-1,j+1}(\mathfrak{L});\ \text{ for } 1 \leq i \leq j \leq r-2 \tag{38}$$

$$a_{i,j} \leq m_{r-j+1} + s_{i,j+1}(\mathfrak{L}) - 2\bar{s}_{i,j}(\mathfrak{L}) + \bar{s}_{i,j-1}(\mathfrak{L});\ \text{ for } 1 \leq i \leq j \leq r-2 \tag{39}$$

$$a_{i,r-1} \leq m_2 + \bar{s}_{i,r-2}(\mathfrak{L}) - 2t_{i-1,r-1}(\mathfrak{L}) \tag{40}$$

$$a_{i,r} \leq m_1 + \bar{s}_{i,r-2}(\mathfrak{L}) - 2t_{i-1,r}(\mathfrak{L}) \tag{41}$$

Let $BZL(\lambda)$ denote the set of Littelmann patterns $\mathfrak{L}$ that are in bijection with elements of the highest weight crystal $\mathcal{C}_\lambda$.

4 The Constructions

We are ready to give the constructions of p parts of WMDS and metaplectic Whittaker functions, i.e., the objects from Sect. 2.4. To emphasize the combinatorial nature of the constructions in this section, we restrict our attention to constructing a polynomial P_λ. The meaning of this polynomial in each of the types was given in Sects. 1.2 and 2.4.

We shall give a polynomial P for any Cartan type in the infinite families A_r, B_r, C_r, and D_r.

The constructions are analogous in different types. Before giving the type-by-type constructions in Sect. 4.3, we begin by summarizing the common elements. In this section and afterward, we shall identify a crystal element with the corresponding Littelmann pattern.

4.1 The Contribution of a Pattern

In all cases, $P = P_\lambda$ is a sum over a crystal of highest weight λ. Using the bijection $b \mapsto \mathfrak{L}(b)$ above, we write P as a sum over $BZL(\lambda)$. We shall have a sum:

$$P = \sum_{\mathfrak{L} \in BZL(\lambda)} G(\mathfrak{L}) \cdot \mathbf{x}^{\mathrm{wt}(\mathfrak{L})} \tag{42}$$

where recall that $\mathbb{C}[\Lambda]$ was identified with a polynomial ring $\mathbb{C}[\mathbf{x}]$. Here $wt(\mathfrak{L})$ is essentially the weight of the pattern given in Sect. 3.4.

The coefficient $G(\mathfrak{L})$ shall be given as a product:

$$G(\mathfrak{L}) = \prod_{i,j} g_{i,j}(\mathfrak{L}) \tag{43}$$

This product is over elements of the Littelmann pattern, and the factor $g_{i,j}(\mathfrak{L})$ depends only on the *decoration* of the element $a_{i,j}$ in $\mathfrak{L}$. For each of the infinite families, we decorate the elements of the pattern $\mathfrak{L}$ according to a *circling* and a *boxing* rule. An entry $a_{i,j}$ may be circled, boxed, neither, or both. Before giving the rules for decorating elements of $\mathfrak{L}$ in Sect. 4.2, we preview how the decorations affect the contribution of $\mathfrak{L}$.

In the case of the constructions in type A, B, and C, the factor $g_{i,j}(\mathfrak{L})$ depends only on the decoration of $a_{i,j}$ and the integer $a_{i,j}$ itself.[1] In particular (in type A and C)[2]:

$$\text{If } a_{i,j} \text{ is circled, then } g_{i,j}(\mathfrak{L}) := \begin{cases} q^{a_{i,j}} & \text{if } a_{i,j} \text{ is not boxed} \\ 0 & \text{if } a_{i,j} \text{ is boxed} \end{cases} \tag{44}$$

If $a_{i,j}$ is *not* circled, then the value of $g_{i,j}(\mathfrak{L})$ is a Gauss sum (see Sect. 2.2).

4.2 Circling and Boxing Rules

Recall that in Sect. 2.3.2, we identified elements of the highest weight crystal $\mathcal{C}_\lambda$ with integral points of the Littelmann polytope. In Sect. 3, we gave a set of constraints that a pattern $\mathfrak{L}(b) \in BZL(\lambda)$ satisfies if it corresponds to an element $b \in \mathcal{C}_\lambda$. The constraints came in the form of inequalities (27)–(30) (these guarantee that b is in the Littelmann cone) and (34)–(38) (these are specific to the polytope and depend on λ). The decoration of an entry $a_{i,j}$ of $\mathfrak{L}$ depends on whether the inequalities involving $a_{i,j}$ are satisfied with an equality.

4.2.1 Circling Rule

The inequalities (27)–(30) involve a single row of the Littelmann pattern. An element $a_{i,j}$ appears in one of these, and that has a lower bound for $a_{i,j}$. The lower bound is of the form:

$$a_{i,j} \geq \begin{cases} \max(a_{i,j+1}, a_{i,j+2}) & \text{if } \mathfrak{L} \text{ is of type } D \text{ and } j = r-2 \\ a_{i,j+2} & \text{if } \mathfrak{L} \text{ is of type } D \text{ and } j = r-1 \\ \frac{1}{2} \cdot a_{i,j+1} & \text{if } \mathfrak{L} \text{ is of type } B \text{ and } j = r-1 \\ 2 \cdot a_{i,j+1} & \text{if } \mathfrak{L} \text{ is of type } B \text{ and } j = r \\ a_{i,j+1} & \text{otherwise} \end{cases} \tag{45}$$

Then $a_{i,j}$ is *circled* if this lower bound (45) holds with an equality.

[1] In type D, the picture is more complex. An analogous construction gives a result slightly different from the one expected from the p-part. The phenomenon of *(symmetric) multiple leaners* accounts for this discrepancy; see Sect. 4.3.4 for details.

[2] Note that in type B, we have the coefficient 1 instead of $q^{a_{i,j}}$. This discrepancy by a factor of q also present when comparing (47) with (46) or (48) can be eliminated with a change of variables in the polynomial $P_\lambda(\mathbf{x})$.

4.2.2 Boxing Rule

The inequalities (34)–(38) have the property that every $a_{i,j}$ entry of a Littelmann pattern appears on the left-hand side of exactly one of them. Let $a_{i,j}$ be *boxed* if this inequality holds with an equality.

4.2.3 Interpretation of Decorations

Observe that an entry is circled or boxed when an inequality defining the Littelmann polytope is satisfied with an equality. This means that the element of the polytope corresponding to $\mathfrak{L}$ is on one of the hyperplanes defining the polytope.

4.3 Constructions Type by Type

We are ready to finish describing the constructions in type A_r, B_r, C_r, and D_r.

4.3.1 Type A

We recall the definition of a p-part from [14]. First recall that the weight $\mathfrak{s}(\mathfrak{L})$ of a pattern was defined in (23) as the sum of entries in columns. In this case, we are interested in the p-part (13); we set $P = P_\lambda(\mathbf{x})$ equal to this p-part with $\lambda = \sum_{i=1}^{r}(l_i + 1)\varpi_i$ and we assume that λ is a strongly dominant weight ($m_i = l_i \geq 1$ for every $1 \leq i \leq r$).

Then P is given by (42) where $\mathrm{wt}(\mathfrak{L}) = \mathfrak{s}(\mathfrak{L})$ and $G(\mathfrak{L})$ is a product as in (43) where the factors of the coefficient are given as follows [14, Chapter 1]:

$$g_{i,j}(\mathfrak{L}) = \begin{cases} 0 & \text{if } a_{i,j} \text{ is circled and boxed} \\ q^{a_{i,j}} & \text{if } a_{i,j} \text{ is circled but not boxed} \\ g(a_{i,j}) & \text{if } a_{i,j} \text{ is boxed but not circled} \\ h(a_{i,j}) & \text{if } a_{i,j} \text{ is neither circled nor boxed} \end{cases} \tag{46}$$

Here $g = g_1$ and $h = h_1$ are the Gauss sums from Sect. 2.2 and q is the order of a residue field.

Remark. *This is the construction of the p part H_Γ from [14].*

4.3.2 Type B

Recall from Sects. 1.2 and 2.4.2 that [16] gives a conjectural formula for a Whittaker function in type B. In this case, we have that $P_\lambda(\mathbf{x})$ is the value $\mathcal{W}(\mathbf{x}, \lambda)$ of a Whittaker function on a torus element.

In this case, let $\mathrm{wt}(\mathfrak{L}) := \sum_{k=1}^{r} s_k \alpha_k - \lambda$ where $\mathfrak{s}(\mathfrak{L}) = (s_1, \ldots, s_r)$ as defined in (24) and (25). Then [16, Conjecture 2] states that P is given by (42) and (43) where the factors are given as follows:

$$g_{i,j}(\mathfrak{L}) = \begin{cases} 0 & \text{if } a_{i,j} \text{ is circled and boxed} \\ 1 & \text{if } a_{i,j} \text{ is circled but not boxed} \\ q^{-a_{i,j}} g_t(a_{i,j}) & \text{if } a_{i,j} \text{ is boxed but not circled} \\ q^{-a_{i,j}} h_t(a_{i,j}) & \text{if } a_{i,j} \text{ is neither circled nor boxed} \end{cases} \tag{47}$$

where g_t and h_t are as in Sect. 2.2 and the subscript t is 1 if $j = r$ and $t = 2$ otherwise.

4.3.3 Type C

Once again we have $P_\lambda(\mathbf{x})$ be the p-part of a Multiple Dirichlet series from [2] or [26]. Let us once again write $\mathrm{wt}(\mathfrak{L}) = \mathfrak{s}(\mathfrak{L})$. To specify P by (42) and (43), we must again specify the factors $g_{i,j}(\mathfrak{L})$:

$$g_{i,j}(\mathfrak{L}) = \begin{cases} q^{a_{i,j}} & \text{if } a_{i,j} \text{ is circled but not boxed} \\ g_t(a_{i,j}) & \text{if } a_{i,j} \text{ is boxed but not circled} \\ h_1(a_{i,j}) & \text{if } a_{i,j} \text{ is neither circled nor boxed and } n | a_{i,j} \\ 0 & \text{otherwise} \end{cases} \tag{48}$$

where again $t = 1$ if $j \neq r$ and $t = 2$ is $j = r$. (We note that [2, (31)] contained a typo that was fixed by [26, (34)]: note that by (Sect. 2.2) if $a_{i,j}$ is neither circled nor boxed, then $g_{i,j}(\mathfrak{L}) = 0$ unless $n|a_{i,j}$.)

4.3.4 Type D

Finally, we recall [22, Conjecture 1], a conjectural expression for the p-part $P_\lambda(\mathbf{x})$ of a Multiple Dirichlet series. In this case, the construction for $P (= P_\lambda)$ is slightly different. Once again, it is a sum (42) over contributions from Littelmann patterns $\mathfrak{L} \in BZL(\lambda)$ and the weight of a pattern is $\mathrm{wt}(\mathfrak{L}) = \mathfrak{s}(\mathfrak{L})$. The contribution $G(\mathfrak{L})$ of a pattern is again a product. However, in this case the factor $g_{i,j}$ is dependent on more than the decoration of $a_{i,j}$.

The definition of $G(\mathfrak{L})$ in [22] is written as a product over connected components of the *decorated graph* $\Gamma(\mathfrak{L})$ of the pattern $\mathfrak{L}$. To give the conjectural construction of the p-part, we introduce some terminology.

The vertices of $\Gamma(\mathfrak{L})$ are the entries of $\mathfrak{L}$. Two entries belong to the same connected component in $\Gamma(\mathfrak{L})$ if they are comparable in the inequalities (30) and they are equal. By the rightmost element of a component C, we mean the entry in C that is positioned rightmost in the Littelmann pattern. A connected component C is called a *multiple leaner* (m.l.) if it consists of entries $a_{i,j_1} = a_{i,j_1+1} = \cdots = a_{i,j_2}$ where $j_1 \leq r-2$, $r+1 \geq j_2$ and $a_{i,j_1-1} > a_{i,j_1}$, $a_{i,j_2} > a_{i,j_2+1}$. By the legs of C we mean the entries $a_{i,j_1} = \cdots = a_{i,r-2}$ and $a_{i,r+1} = \cdots = a_{i,j_2}$; the entry on the endpoint of the shorter leg of Γ is a_{i,j_1} or a_{i,j_2}. The component C is called a *symmetric multiple leaner* (s.m.l.) if in addition $j_2 = \bar{j}_1$; in this case, we define its length to be $l(C) = r - j_1$, half the number of its vertices.

We may write $G(\mathfrak{L})$ as in (43) but to define $g_{i,j}(\mathfrak{L})$, we write [22, 5.5]

$$\sigma(C) = \prod_{a_{i,j} \in C} g_{i,j}(\mathfrak{L}) \tag{49}$$

and give $\sigma(C)$ in terms of standard contributions of the entries it contains. Let [22, 5.5]

$$\sigma(y) = \begin{cases} 0 & \text{if the entry } a \text{ is circled and boxed} \\ h_1(a) \cdot q^{-a} & \text{if the entry } a \text{ is not boxed and not circled} \\ g_1(a) \cdot q^{-a} & \text{if the entry } a \text{ is boxed and not circled} \end{cases} \tag{50}$$

We then define $\sigma(C)$ to be as follows:

- $\sigma(C) = 0$ if any $a_{i,j} \in C$ is both circled and boxed.
- $\sigma(C) = \sigma(a)$ if C is not a m.l. and a is its rightmost element, or C is a m.l. that is not symmetric and a is the endpoint of its shorter leg.
- $\sigma(C) = \sigma(a)(1 - q^{-l(C)})$ if C is a s.m.l., $a \neq 0$ is its rightmost element, and a is unboxed.
- $\sigma(C) = \sigma(a_{i,j})\sigma(a_{i,j-1})q^{1-l(C)}$ if C is a s.m.l., $a_{i,j}$ is its rightmost element, and $a_{i,j} \neq 0$ is boxed.
- $\sigma(C) = 1$ if C is a s.m.l. with zero entries.

Note that $\sigma(C)$ is a product over $g_{i,j}(\mathfrak{L})$, but now $g_{i,j}(\mathfrak{L})$ depends not only on the decoration of $a_{i,j} \in \mathfrak{L}$, but also on the position of $a_{i,j}$ within a connected component of $\Gamma(\mathfrak{L})$, and whether that component is a (symmetric) multiple leaner or not.

5 Branching

In the previous section, we described constructions of polynomials $P_\lambda(\mathbf{x})$ that are of interest from a number-theoretic perspective as explained in Sect. 2.4. We also mentioned in Sect. 1.4 that elucidating the relationship of the constructions with the branching properties of highest weight crystals can be the key to understanding some of their properties. In this section, we take a closer look at how the branching properties of crystals manifests in these constructions.

In all of the examples above, the polynomial P was associated to a crystal $\mathcal{C}_\lambda$ corresponding to a root system Φ or rank r, with Cartan type in one of the infinite families A_r, B_r, C_r, or D_r. When the edges of $\mathcal{C}_\lambda$ labeled by α_r are omitted, the remaining graph is a disjoint union of rank $r-1$ crystals of the same Cartan type, but rank $r-1$:

$$\mathcal{C}_\lambda = \bigsqcup_\mu \mathcal{C}_\mu \tag{51}$$

We now wish explain how P_λ can be written in terms of the polynomials P_μ associated to those crystals.

The construction of P_λ is given (cf. (42)) as a sum over Littelmann patterns $\mathfrak{L} \in BZL(\lambda)$, each contributing a term $G(\mathfrak{L})\mathbf{x}^{\mathrm{wt}(\mathfrak{L})}$. We examine how the sets $BZL(\mu)$ corresponding to μ in (51) can be recovered from $BZL(\lambda)$ by giving a pattern $\mathfrak{L}' \in BZL(\mu)$ for any pattern $\mathfrak{L} \in BZL(\lambda)$ in Sect. 5.1. We indicate a method of computing the weights μ that appear in (51) in Sect. 5.2. Then in Sects. 5.2 and 5.3, we explain how for such pairs $\mathfrak{L}$ and $\mathfrak{L}'$ the contributions $G(\mathfrak{L})$ and $\mathrm{wt}(\mathfrak{L})$ can be written in terms of $G(\mathfrak{L}')$, $\mathrm{wt}(\mathfrak{L}')$ and μ. For the remainder of the discussion, let us fix a dominant weight $\lambda = \sum_{k=1}^r m_k \varpi_k$ and the corresponding crystal $\mathcal{C}_\lambda$.

5.1 Patterns with Fixed Top Row

Recall the bijection between elements of the crystal $\mathcal{C}_\lambda$ and Littelmann patterns given in Sect. 3.3. For an element $b \in \mathcal{C}_\lambda$, the entries of the pattern $\mathfrak{L}(b) = \mathfrak{L}_\lambda(b)$ are the lengths of the segments in the BZL path of b. Here $BZL(b)$ corresponds to the choice of a particular long word $\underline{w}_0$.

Let the element $b \in \mathcal{C}_\lambda$ belong to $\mathcal{C}_\mu$ in the decomposition (51). We shall sometimes write b' when we mean b as an element of the abstract crystal $\mathcal{C}_\mu$; write $\mathfrak{L}' = \mathfrak{L}(b') \in BZL(\mu)$.

As remarked in Sect. 3.1, the choices made in (15) all have the property that the long word $\underline{w}_0^{X_r}$ chosen in rank r starts with the long word $\underline{w}_0^{X_{r-1}}$ chosen in rank $r-1$. Together with the shape of the patterns (see Sect. 3.2), this means that $\mathfrak{L}(b')$ is the same as $\mathfrak{L}(b)$ without its first row. This argument proves the following.

Lemma. *Let $b_\mu \in \mathcal{C}_\mu \subset \mathcal{C}_\lambda$ be the highest element within a crystal in the decomposition (51). Let $\mathfrak{L}_\mu(b)$ and $\mathfrak{L}_\lambda(b)$ denote the Littelmann patterns corresponding to any $b \in \mathcal{C}_\mu \subset \mathcal{C}_\lambda$ as an element of $\mathcal{C}_\mu$ and $\mathcal{C}_\lambda$, respectively. Then for any $b \in \mathcal{C}_\mu \subset \mathcal{C}_\lambda$ the top row of $\mathfrak{L}_\lambda(b)$ is the same as the top row of $\mathfrak{L}_\lambda(b_\mu)$, and $\mathfrak{L}_\mu(b)$ can be recovered from $\mathfrak{L}_\lambda(b)$ by deleting the top row.*

5.2 The Weights in the Decomposition

In light of Lemma 5.1 and the inequalities on the top row of $\mathfrak{L} \in BZL(\lambda)$, we can describe the highest weights μ appearing in the decomposition (51) by computing the weight of the highest element $b_\mu \in \mathcal{C}_\mu$. Recall that by (22) $\lambda - \mathrm{wt}(b_\mu) = \lambda - \mu$ can be expressed in terms of $\mathfrak{s}(\mathfrak{L}_\lambda(b_\mu))$:

$$\lambda - \mu = \sum_{k=1}^{r} \mathfrak{s}_k(\mathfrak{L}_\lambda(b_\mu)) \cdot \alpha_k \tag{52}$$

and furthermore that we have:

$$\mathfrak{s}(\mathfrak{L}_\lambda(b)) = \mathfrak{s}(\mathfrak{L}_\lambda(b_\mu)) + \mathfrak{s}(\mathfrak{L}_\mu(b)) \text{ for any } b \in \mathcal{C}_\mu \subset \mathcal{C}_\lambda \tag{53}$$

Note that since the entries under the first row of $\mathfrak{L}_\lambda(b_\mu)$ are all zero, the right-hand side of (52) can be written entirely in terms of the entries in the first row of $\mathfrak{L}_\lambda(b_\mu)$. The inequalities restricting the first row of a $\mathfrak{L} \in BZL(\lambda)$ involve no entries from any other row (cf. Sect. 3.5). It follows that given a highest weight λ, we can recover the set of weights μ that appear in the decomposition (51). (This involves expressing the simple roots α_k in terms of the fundamental weights, and carefully examining the restrictions on entries of the first row of a pattern $\mathfrak{L} \in BZL(\lambda)$.) We omit further discussion of this here and refer the reader to [14, (2.4)] for an example of a similar statement in Cartan type A.

5.3 Branching and Contributions

Let b be an element $b \in \mathcal{C}_\mu \subset \mathcal{C}_\lambda$. Let $\mathfrak{L} = \mathfrak{L}_\lambda(b)$ and $\mathfrak{L}' = \mathfrak{L}_\mu(b)$. We wish to write $P_\lambda(\mathbf{x})$ as a sum

$$P_\lambda(\mathbf{x}) = \sum_{\mu} p(\mu) \cdot P_\mu(\mathbf{x}) \tag{54}$$

where the weights μ are the ones of the decomposition (51) and $p(\mu)$ is a monomial.

Before we explain why this decomposition is possible for the polynomials $P_\lambda(\mathbf{x})$ we remark on the terminology of "branching." Recall that the polynomial $P_\lambda(\mathbf{x})$ can be thought of as a deformation of a highest weight character. Equation (54) has a clear analogue for highest weight characters. Let us write V_λ and V_μ for the irreducible representations of highest weight λ and μ (in rank r and $r-1$), respectively. If $P_\lambda(\mathbf{x})$ were the character of V_λ, then the coefficient in the monomial $p(\mu)$ would match the multiplicity of the V_μ in the restriction of V_λ to a subalgebra of corank one determined by the first $r-1$ simple roots.

The contribution of a pattern $\mathfrak{L}$ is of the form $G(\mathfrak{L})\mathbf{x}^{\mathrm{wt}(\mathfrak{L})}$ as in (42). The term $\mathbf{x}^{\mathrm{wt}(\mathfrak{L})}$ depends only on $\mathfrak{s}(\mathfrak{L})$. It follows from (53) that the first $r-1$ component of $\mathfrak{s}(\mathfrak{L})$ is the tuple $\mathfrak{s}(\mathfrak{L}')$, while $\mathfrak{s}_k(\mathfrak{L}) = \mathfrak{s}_k(\mathfrak{L}(b_\mu))$ depends only on μ.

We turn next to the coefficient $G(\mathfrak{L})$. Recall from Sect. 4.1 and in particular (43) that $G(\mathfrak{L})$ is a product of factors $g_{i,j}(\mathfrak{L})$. The factor $g_{i,j}(\mathfrak{L})$ essentially depends on the decoration of an entry $a_{i,j}$ in $\mathfrak{L}$, and the decorations in turn depend on whether the inequalities imposed on the entry $a_{i,j}$ by $\mathfrak{L}$ being an element of $BZL(\lambda)$ are satisfied strictly or with an equality. It is immediate that the factors $g_{1,j}(\mathfrak{L})$ corresponding to entries of the first row of $\mathfrak{L}$ depend only on μ. Closer examination of the inequalities imposed on the lower rows and the decorations show that in fact, $g_{i,j}(\mathfrak{L})$ can be written as a product of $g_{i-1,j-1}(\mathfrak{L}')$ and a factor that depends only on μ, and not the element $b \in \mathcal{C}_\mu$.

Thus we may conclude that one may in fact decompose $P_\lambda(\mathbf{x})$ as in (54). For a precise statement of this flavor in Cartan type A, see [44, Proposition 16.].

Acknowledgements I would like to thank the editors of this volume for giving me the opportunity to contribute. I am grateful to several people for helpful conversations and advice during the writing of this chapter, including Holley Friedlander, Paul E Gunnells, Dinakar Muthiah, and Manish Patnaik. During parts of the writing process, I was a postdoctoral fellow at the University of Alberta and a visiting assistant professor at the University of Massachusetts, Amherst, and I am grateful to both institutions. While at the University of Alberta, I was supported through Manish Patnaik's Subbarao Professorship in number theory and an NSERC Discovery Grant. I also thank the referees for their insightful comments for the improvement of this chapter. In particular, I thank one of the referees for their comments on the connections to character theory and on Gauss sums.

References

1. J. Beineke, B. Brubaker, S. Frechette, Weyl group multiple Dirichlet series of type C. Pac. J. Math. **254**(1), 11–46 (2011). https://doi.org/10.2140/pjm.2011.254.11. MR2900659
2. J. Beineke, B. Brubaker, S. Frechette, A crystal definition for symplectic multiple Dirichlet series, *Multiple Dirichlet series, L-functions and Automorphic Forms* (Springer, Berlin, 2012), pp. 37–63
3. A. Berenstein, A. Zelevinsky, String bases for quantum groups of type A_r. Adv. Sov. Math. **16**(1), 51–89 (1993)
4. A. Berenstein, A. Zelevinsky, Tensor product multiplicities, canonical bases and totally positive varieties. Invent. Math. **143**(1), 77–128 (2001). https://doi.org/10.1007/s002220000102. MR1802793
5. A. Berenstein, A. Zelevinsky et al., Canonical bases for the quantum group of type A_r and piecewise-linear combinatorics. Duke Math. J. **82**(3), 473–502 (1996)
6. N. Bourbaki, *Lie Groups and Lie Algebras, Chapters 4–6*, Elements of Mathematics (Springer, Berlin, 2002). Translated from the 1968 French original by Andrew Pressley. MR1890629 (2003a:17001)
7. A. Braverman, D. Gaitsgory et al., Crystals via the affine Grassmannian. Duke Math. J. **107**(3), 561–575 (2001)
8. B. Brubaker, S. Friedberg, Whittaker coefficients of metaplectic Eisenstein series. Geom. Funct. Anal. **25**(4), 1180–1239 (2015)
9. B. Brubaker, A. Schultz, On Hamiltonians for six-vertex models (2016), arXiv:1606.00020

10. B. Brubaker, V. Buciumas, D. Bump, A Yang-Baxter equation for metaplectic ice (2016), arXiv:1604.02206
11. B. Brubaker, D. Bump, S. Friedberg, Weyl group multiple Dirichlet series II: the stable case. Invent. Math. **165**(2), 325–355 (2006). MR2231959 (2007g:11056)
12. B. Brubaker, D. Bump, G. Chinta, S. Friedberg, J. Hoffstein, Weyl group multiple Dirichlet series I, in *Multiple Dirichlet Series, Automorphic Forms, and Analytic Number Theory* (2006), pp. 91–114
13. B. Brubaker, D. Bump, G. Chinta, S. Friedberg, P.E. Gunnells, Metaplectic ice (2010), arXiv:1009.1741
14. B. Brubaker, D. Bump, S. Friedberg, *Weyl Group Multiple Dirichlet Series: Type A Combinatorial Theory*, Annals of Mathematics Studies (Princeton University Press, Princeton, 2011). MR2791904
15. B. Brubaker, D. Bump, S. Friedberg, Weyl group multiple Dirichlet series, Eisenstein series and crystal bases. Ann. Math. **173**(2), 1081–1120 (2011). https://doi.org/10.4007/annals.2011.173.2.13. MR2776371
16. B. Brubaker, D. Bump, G. Chinta, P.E. Gunnells, Metaplectic Whittaker functions and crystals of type B, in *Multiple Dirichlet Series, L-functions and Automorphic Forms* (2012), pp. 93–118
17. D. Bump, Introduction: multiple Dirichlet series, in *Multiple Dirichlet Series, L-functions and Automorphic Forms* (2012), pp. 1–36
18. D. Bump, A. Schilling, *Crystal Bases: Representations and Combinatorics* (World Scientific Publishing Company, Singapore, 2017)
19. W. Casselman, J. Shalika, The unramified principal series of p-adic groups II: the Whittaker function. Compos. Math. **41**(2), 207–231 (1980). MR581582 (83i:22027)
20. G. Chinta, S. Friedberg, J. Hoffstein, Multiple Dirichlet series and automorphic forms. Proc. Symp. Pure Math. **75**, 3–42 (2006)
21. G. Chinta, P.E. Gunnells, Constructing Weyl group multiple Dirichlet series. J. Am. Math. Soc. **23**(1), 189–215 (2010). https://doi.org/10.1090/S0894-0347-09-00641-9. MR2552251 (2011g:11100)
22. G. Chinta, P.E. Gunnells, Littelmann patterns and Weyl group multiple Dirichlet series of type D, *Multiple Dirichlet Series, L-functions and Automorphic Forms* (Springer, Berlin, 2012), pp. 119–130
23. G. Chinta, P.E. Gunnells, A. Puskás, Metaplectic Demazure, operators and Whittaker functions. Indiana Univ. Math. J. **66**(3) (2014) arXiv:1408.5394
24. G. Chinta, O. Offen, A metaplectic Casselman-Shalika formula for GL_r. Am. J. Math. **135**(2), 403–441 (2013)
25. H. Davenport, H. Hasse, Die Nullstellen der Kongruenzzetafunktionen in gewissen zyklischen Fällen. J. Reine Angew. Math. **172**(151–182), 198–201 (1934)
26. S. Friedberg, L. Zhang, Eisenstein series on covers of odd orthogonal groups. Am. J. Math. **137**(4), 953–1011 (2015)
27. H. Friedlander, L. Gaudet, P.E. Gunnells, Crystal graphs, Tokuyama's theorem, and the Gindikin-Karpelevič formula for G2. J. Algebr. Comb. **41**(4), 1089–1102 (2015). https://doi.org/10.1007/s10801-014-0567-9. MR3342714
28. A.M. Hamel, R.C. King, Half-turn symmetric alternating sign matrices and Tokuyama type factorisation for orthogonal group characters. J. Combin. Theory Ser. A **131**, 1–31 (2015). https://doi.org/10.1016/j.jcta.2014.11.005. MR3291473
29. J. Kamnitzer, Mirković-Vilonen cycles and polytopes. Ann. Math. **171**, 245–294 (2010)
30. M. Kashiwara, *Representations of Groups, (Banff, AB, 1994)*, vol. 16 (1995), pp. 155–197
31. M. Kashiwara, T. Nakashima, Crystal graphs for representations of the q-analogue of classical Lie algebras. J. Algebra **165**(2), 295–345 (1994). MR1273277
32. V. Lakshmibai, C.S. Seshadri, Standard monomial theory, in *Proceedings of the Hyderabad Conference on Algebraic Groups, Hyderabad, 1989* (1991), pp. 279–322. MR1131317
33. P. Littelmann, A Littlewood-Richardson rule for symmetrizable Kac-Moody algebras. Invent. Math. **116**(1), 329–346 (1994)
34. P. Littelmann, Cones, crystals, and patterns. Transform. Groups **3**(2), 145–179 (1998)

35. G. Lusztig, Canonical bases arising from quantized enveloping algebras. J. Am. Math. Soc. **3**(2), 447–498 (1990)
36. G. Lusztig, Canonical bases arising from quantized enveloping algebras II. Prog. Theor. Phys. Suppl. **102**, 175–201 (1991)
37. G. Lusztig, An algebraic-geometric parametrization of the canonical basis. Adv. Math. **120**(1), 173–190 (1996)
38. H. Matsumoto, Sur les sous-groupes arithmétiques des groupes semi-simples déployés. Ann. Sci. École Norm. Sup. **2**(1), 1–62 (1969). MR0240214
39. J.P. McNamara, Metaplectic Whittaker functions and crystal bases. Duke Math. J. **156**(1), 1–31 (2011). https://doi.org/10.1215/00127094-2010-064. MR2746386
40. P.J. McNamara, The metaplectic Casselman-Shalika formula. Trans. Am. Math. Soc. **368**(4), 2913–2937 (2016)
41. J. Neukirch, *Algebraic Number Theory*, vol. 322 (Springer Science and Business Media, Berlin, 2013)
42. M. Patnaik, A. Puskás, Metaplectic covers of Kac-Moody groups and Whittaker functions (2017), arXiv:1703.05265
43. M. Patnaik, A. Puskás, On Iwahori-Whittaker functions for metaplectic groups. Adv. Math. **313**, 875–914 (2017)
44. A. Puskás, Whittaker functions on metaplectic covers of GL(r) (2016), arXiv:1605.05400
45. T. Tokuyama, A generating function of strict Gel′fand patterns and some formulas on characters of general linear groups. J. Math. Soc. Jpn. **40**(4), 671–685 (1988). https://doi.org/10.2969/jmsj/04040671. MR959093 (89m:22019)
46. K. Yamamoto, On a conjecture of Hasse concerning multiplicative relations of Gaussian sums. J. Comb. Theory **1**(4), 476–489 (1966)